

VISUAL ANALYTIC OF UMP ATTRITION

NUR ALNISA' ANIS ALANNA RUZELAN

Bachelor of Computer Science
(Graphic & Multimedia Technology)

UNIVERSITI MALAYSIA PAHANG



SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Computer Science (Graphic & Multimedia Technology) with honors.

(Supervisor's Signature)

Full Name :

Position :

Date :



STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

(Student's Signature)

Full Name : NUR ALNISA' ANIS ALANNA BINTI RUZELAN

ID Number : CD15046

Date : January 2019

VISUAL ANALYTICS OF UMP ATTRITION

NUR ALNISA' ANIS ALANNA BINTI RUZELAN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science(Graphic & Multimedia Technology) with honors

Faculty of Computer System & Software Engineering
UNIVERSITI MALAYSIA PAHANG

JANUARY 2019

ACKNOWLEDGEMENTS

Bismillahirrahmanirrahim,

Thanks to Allah SWT, whom with His willing giving me the opportunity to complete this Final Year Report which is Visual Analytics of UMP Attrition. This final year project was prepared for Faculty of Computer Systems & Software Engineering (FSKKP), University Malaysia Pahang (UMP) is basically for student in final year to complete the undergraduate program that leads to the degree of Bachelor of Computer Science (Graphic & Multimedia Technology).

Firstly, I would like to express my deepest thanks to Ts. En. Aziman Abdullah, a lecturer at FSKKP and also assign as my supervisor who had guided me a lot during two semester session 2017/2018. I also want thanks staffs of Jabatan Hal Ehwal Pelajar(JHEPA) and Bahagian Pengurusan Akademik (BPA) UMP for their cooperation during completing my final year project that had given valuable information, suggestion and guidance.

Deepest thanks and appreciation to my parents for their moral support, encouragement that give me strength throughout this final year project and provided good opportunities for a good education.

Last but not least, I would like to thanks to all my friends and everyone for their encouragement cooperation, encouragement, constructive suggestion and full of support for the report completion, from the beginning till the end that have been contributed by supporting my work and help myself during the final year project progress till it fully completed.

ABSTRAK

Peningkatan pendaftaran pelajar di Universiti Malaysia Pahang (UMP) dari tahun ke tahun disebabkan lebih banyak program yang ditawarkan telah menyebabkan penambahan jumlah pelajar berhenti belajar tanpa bergraduasi. Untuk menyaht Kementerian Pelajaran Tinggi Malaysia, UMP perlu menyediakan lebih banyak akses kepada pendidikan tinggi dan program baru yang perlu ditawarkan kepada pelajar untuk mencapai tahap ekonomi digital baru. Bagi melaksanakan perkara tersebut, pendekatan yang lebih bijak dengan menggunakan data analisis adalah menjadi satu keperluan pada masa kini dalam mengendalikan kes pelajar yang berhenti belajar. Oleh itu, kajian ini bertujuan untuk mengkaji faktor akademik berdasarkan Program Learning Outcomes (PLO) yang sangat mempengaruhi atau menyebabkan berlakunya kes pelajar berhenti belajar. Dengan menggunakan kaedah regrasi, analisis di antara Purata Nilai Gred Kumulatif (CGPA) dan corak prestasinya terhadap lapan hasil pembelajaran program diukur dengan menggunakan kaedah Pearson's khususnya kepada para pelajar yang berisiko gagal dalam pembelajaran seterusnya berhenti belajar. Nilai dari semua lapan PLO diseragamkan dengan skala standard antara 0-1 dengan mengira jumlah markah yang diberikan terhadap setiap hasil pembelajaran program. Hasil daripada analisis menunjukkan terdapat hubungan yang positif antara CGPA dan PLO. Kajian ini menunjukkan bahawa terdapat kemungkinan pelajar berisiko boleh dipantau dengan betul dengan membangunkan model dengan meramalkan pelajar yang berisiko untuk berhenti belajar berdasarkan prestasi hasil pembelajaran program.

ABSTRACT

The increasing pattern of student enrollment at University Malaysia Pahang (UMP) due to more programs have been offered as well as increasing number of causing the increasing number of students attrition. To respond with Malaysia Higher Education Blueprint, UMP need to provide more access to higher education and more new programs need to be offered to embrace a new digital economy. To do so, smarter approach with the use of data analytics is a must in managing student attrition cases. Therefore, this study aims to investigate the academic factors based on the standard Program Learning Outcomes (PLO) that highly influence or affecting the attrition cases. By using simple linear regression, analysis of the correlation between Cumulative Grade Point Average (CGPA) and its performance pattern against eight program learning outcomes were measured by calculating Pearson correlation coefficient of the attrition case specifically of the dropout student. The value of all eight program outcomes were standardized with standard scale between 0-1 by calculating the ration between the total marks awarded against the total of full mark of each program learning outcomes. The findings from the analysis show there are strong positive correlation between CGPA and PLO. This study shows there is possibility how risky students can be properly monitored by developing predictive model based on program learning outcomes performances.

TABLE OF CONTENT

DECLARATION	
TITLE PAGE	
ACKNOWLEDGEMENTS	ii
ABSTRAK	iii
ABSTRACT	iv
TABLE OF CONTENT	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1 INTRODUCTION	1
1.1 Background of study	1
1.2 Problem Statement	3
1.3 Research Question	3
1.4 Research Objectives	3
1.5 Research Hypothesis	4
1.6 Scope	4
1.7 Significance	4
1.8 Thesis Organization	5
CHAPTER 2 LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Alumni Analytics	6

2.2.1	Dropout Issues	6
2.3	Predictive Analytics	7
2.4	Dependent Variable	9
2.4.1	Cumulative Grade Point Average (CGPA)	9
2.5	Independent Variables	9
2.5.1	Program Learning Outcomes (PLO)	9
 CHAPTER 3 METHODOLOGY		 12
3.1	Introduction	12
3.2	Research Methodology	12
3.3	Quantitative Approach	12
3.4	Predictive Analytic Process	13
3.4.1	Project Definition	13
3.4.2	Data Collection	14
3.4.3	Data Processing	14
3.4.4	Data Cleaning	15
3.4.5	Exploratory Data Analysis	15
3.4.6	Model Development	16
3.4.7	Model Refinement	16
3.4.8	Insights and Reporting	16
3.5	Analytical Technique	17
3.5.1	Simple Linear Regression	17
3.6	Hardware and Software	18
 CHAPTER 4 RESULTS AND DISCUSSION		 19
4.1	Introduction	19

4.2	Discussion of Findings	19
4.2.1	Demographic Findings	19
4.3	Model Testing	20
4.3.1	Predictive model	21
4.3.2	Relationship between CGPA and PLO	22
4.3.3	Ranking of influence between PLO	23
4.3.4	Strongest influence factor based on PLO	24
4.3.5	Relationship between PLO	25
CHAPTER 5 CONCLUSION		32
5.1	Introduction	32
5.2	Conclusion	32
5.3	Research Constraint	32
5.4	Future Work	33
REFERENCES		34
APPENDIX A GANTT CHART		36

LIST OF TABLES

Table 2.1	Academic Standing	9
Table 4.2	Demographic Characteristics of Students	20

LIST OF FIGURES

Figure 2.1	Example of PLO in student's ecomm	10
Figure 3.1	The predictive analytics process	13
Figure 3.2	The raw data from data collection	14
Figure 3.3	Data cleaning in PowerBi	15
Figure 3.4	The example of data analysis in visual approach	16
Figure 4.1	The relationship between CGPA and PLO1	21
Figure 4.2	The relationship between CGPA, PLO1 and PLO6	22
Figure 4.3	The ranking of the PLO	23
Figure 4.4	The strongest influence factor based on each PLO	24
Figure 4.5	The relationship between PLO1 and PLO6	25
Figure 4.6	The relationship between PLO2 and PLO5	26
Figure 4.7	The relationship between PLO3 and PLO8	27
Figure 4.8	The relationship between PLO4 and PLO6	28
Figure 4.9	The relationship between PLO5 and PLO6	29
Figure 5.0	The relationship between PLO6 and PLO7	30
Figure 5.1	The relationship between PLO7 and PLO8	31

LIST OF ABBREVIATIONS

CAIC	Centre of Academic Innovation & Competitiveness
CGPA	Cumulative Grade Point Average
FIM	Faculty of Industrial Management
FIST	Faculty of Industrial Science & Technology
FKASA	Faculty of Civil Engineering & Earth Resources
FKEE	Faculty of Electrical & Electronic Engineering
FKKSA	Faculty of Chemical & Natural Resources Engineering
FKM	Faculty of Mechanical Engineering
FKP	Faculty of Manufacturing Engineering
FSKKP	Faculty of Computer Systems & Software Engineering
FTek	Faculty of Engineering Technology
PLO	Program Learning Outcomes

CHAPTER 1

INTRODUCTION

1.1 Background of study

Malaysia has reached a gross higher education enrolment rate of 48% where as 70% increase in enrolment of total higher education since 2004 until 2014 to reach 1.2 million students in public and private universities in a decade (Ministry of Education Malaysia, 2015). The enrolment for undergraduate's students has been increased sixfold and for post graduates students has increased tenfold in Masters and PhD enrolment. With this increase, Malaysia now had been ranks the third among the ASEAN countries in Masters and PhD enrolment which is behind Thailand and Singapore. In order to respond with Malaysia Higher Education Blueprint to increases the access to and enrolment in higher education, University Malaysia Pahang (UMP) need to provide more access to higher education and more new programs need to be offered to embrace a new digital economy. The increasing of the access and enrolment of higher education may concern the retention rate and attrition rate in UMP.

Since, attrition rate at universities has been concern among educators for many decades as the impact of student's attrition is the most significance for students itself, lecturers and universities. Student's retention defined as the students who turn into the same university after completing a semester in the university. The increasing of student's retention is a long term goal for any universities in the world. The higher the student's retention rate, the most likely that the university may position higher in ranking and the secured the government funds which may lead to the easier path for program accreditations. The attrition rate is the rate of dropout among the participants or subjects in a market research. In this study focused on student's attrition rate where refers to rate of dropout among the students in UMP. The students may have highest risk of dropping out at the beginning, middle and also last at their study depend on the various factors. By

early detection about the students with higher risk of being dropping out from the university may improve the attrition rate and also the student's retention rate in the university. Therefore, it is really important to have such a predictive model to make an early prediction about the dropout's students that need to be addressed effectively. Thus, this study clearly need to develop predictive models which is can potently identify the students who has the highest potential to become a dropouts.

Visual analytics is defined as science of analytical reasoning facilitated by interactive visual interfaces(Thomas, 2007). According to (Keim, Mansmann, Stoffel, & Ziegler, 2009; Kohlhammer, Keim, Pohl, Santucci, & Andrienko, 2011), visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets. The goal of the visual analytics is permit person to synthesize information and derive insight from huge, effective, ambiguous and often clash data, detect the expected and determine the unexpected and communicate assessment effectively for action. Visual analytics is not just a visualization. Rather than be seen as decision making, merging visualization, human factors and data analysis, it can be as integral approach.

According to the previous study (Ameri, Fard, Chinnam, & Reddy, 2016), there many explanatory model which help educational institution to predict the dropout's students were found such as traditional methods. For a decade, the traditional methods such as regression and logistic regression had been used to predict the student's dropout. Regression models are widely used to address scientific questions of interest regards to the associations among the sets of variables(Fitzmaurice, 2016). Two most common goals in regression are explanation and prediction. Regression model which used to estimate the effect of an explanatory response is called as explanation while prediction of unobserved responses for a new individuals or prediction of future values of the response based on present values of explanatory variable is called as prediction. Although in the previous study had experiment in with various strategies and method to predict student's dropout but the implementation of analytics especially in predictive analytics had not been really emphasized in Malaysia. Therefore, in this study investigated the implementation of predictive analytics using prediction regression to predict the student's dropout in UMP. Thus, the issues about the predictive analytics will be answer in this paper based on the discussion, result and recommendation.

1.2 Problem Statement

The increasing pattern of students enrolment at University Malaysia Pahang(UMP) due to more programmes have been offered as well as increasing number of student intakes causing an increasing number of student attrition. When there is higher attrition rate among the students, the less likely that the university will position at the higher ranking among the university in Malaysia. Thus, there will cause the increasing feeling pressure to the director and administrators as well as to lecture in the faculty.

Besides, the existing data has constraint in analysing because there is lack of professional staff and time. Thus the existing data cannot give a predictive output to the stakeholders. As the data cannot prevent the student's from being dropout so there will increasing in attrition rate in UMP.

The prediction of student's dropout also lack in research. There is lack of references about the prediction of dropout in higher education and cause the increasing of attrition rate in the university. Thus, the references about the prediction of student's dropout is important as the prevention from dropout reduce the attrition rate in the university.

1.3 Research Question

- What if alumni analytics can provide the prediction about the student's dropout to the education service provider's especially private sector?

1.4 Research Objectives

- i. To analyse the raw data into predictable information for the stakeholders by using predictive analytics.
- ii. To implement the dropout's data into visual analytics approach for decision making.
- iii. To evaluate the effectiveness of data acquisition method and visualization (dashboard).

REFERENCES

- Alharthi, H. (2018). Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *Journal of Infection and Public Health*, 11(6), 749–756.
<https://doi.org/10.1016/j.jiph.2018.02.005>
- Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016). Survival Analysis based Framework for Early Prediction of Student Dropouts. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16*, (October), 903–912. <https://doi.org/10.1145/2983323.2983351>
- Belo, P., & Oliveira, C. (2015). The Relation between Experiences and Expectations with University Dropout. *Procedia - Social and Behavioral Sciences*, 187, 98–101.
<https://doi.org/10.1016/j.sbspro.2015.03.019>
- Bichsel, J. (2012). Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations. *EDUCAUSE: Center for Applied Research*, (August 2012), 1–31.
<https://doi.org/10.13140/RG.2.1.1064.6244>
- Bonneau, K., & Management, D. (2007). Brief 3 : What is a Dropout ?, 14–17.
- Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic Analytics. *Educause Review*.
- De Witte, K., Cabus, S., Thyssen, G., Groot, W., & Van Den Brink, H. M. (2013). A critical review of the literature on school dropout. *Educational Research Review*, 10, 13–28.
<https://doi.org/10.1016/j.edurev.2013.05.002>
- Fitzmaurice, G. M. (2016). Regression. *Diagnostic Histopathology*, 22(7), 271–278.
<https://doi.org/10.1016/j.mpdhp.2016.06.004>
- Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach. Analysis* (Vol. 1).
<https://doi.org/10.2139/ssrn.263511>
- Garrison, D. R. (1987). Researching dropout in distance education. *Distance Education*, 8(1), 95–101. <https://doi.org/10.1080/0158791870080107>
- Keim, D., Mansmann, F., Stoffel, A., & Ziegler, H. (2009). Visual Analytics. *Encyclopedia of Database Systems*.
- Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., & Andrienko, G. (2011). Solving problems with visual analytics. *Procedia Computer Science*, 7, 117–120.
<https://doi.org/10.1016/j.procs.2011.12.035>

- Ministry of Education Malaysia. (2015). Malaysia Education Blueprint 2015-2025. *Ministry of Education Malaysia, 2025*. Retrieved from <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>
- Mohamad, N., Ahmad, N. B., Norhayati, D., & Jawawi, A. (2018). Malaysia MOOC : Improving Low Student Retention with Predictive Analytics, 7, 398–405. <https://doi.org/10.14419/ijet.v7i2.29.13305>
- Nistor, N., & Neubauer, K. (2010). From participation to dropout: Quantitative participation patterns in online university courses. *Computers and Education, 55*(2), 663–672. <https://doi.org/10.1016/j.compedu.2010.02.026>
- Rumberger, R. W., & Lamb, S. P. (2003). The early employment and further education experiences of high school dropouts: A comparative study of the United States and Australia. *Economics of Education Review, 22*(4), 353–366. [https://doi.org/10.1016/S0272-7757\(02\)00038-9](https://doi.org/10.1016/S0272-7757(02)00038-9)
- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology, 98*, 146–151. <https://doi.org/10.1016/j.jclinepi.2017.12.006>
- Thomas, J. J. J. (2007). Visual analytics : why now ? *Information Visualization, 6*(1), 104–106. <https://doi.org/10.1057/palgrave.ivs.9500148>
- Washington, F. (2006). Graduation and Dropout Statistics, (September).
- Williams, C. (2007). Research Methods. *Journal of Business & Economic Research, 5*(3), 65–72. <https://doi.org/10.1093/fampract/cmi221>
- Zheng, J., Tang, J., & Gao, J. (2012). Construction of Graduate Employment Service System Based on Public Information Platform. *AASRI Procedia, 1*, 410–415. <https://doi.org/10.1016/j.aasri.2012.06.064>