MySPPS : THE DEVELOPMENT OF STUDENT'S PERFORMANCE PREDICTION SYSTEM USING MACHINE LEARNING

NURUL HIDAYAH BINTI JEMAT

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

FAKULTI KOMPUTERAN DAN TEKNOLOGI DIGITAL

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2022

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

MySPPS : THE DEVELOPMENT OF STUDENT'S PERFORMANCE PREDICTION SYSTEM USING MACHINE LEARNING

NURUL HIDAYAH BINTI JEMAT

D20192090502

LAPORAN TAHUN AKHIR DIKEMUKAKAN BAGI MEMENUHI SYARAT UNTUK MEMPEROLEH IJAZAH SARJANA MUDA KEJURUTERAAN PERISIAN (PERISIAN PENDIDIKAN) DENGAN KEPUJIAN

FAKULTI KOMPUTERAN DAN TEKNOLOGI DIGITAL

UNIVERSITI PENDIDIKAN SULTAN IDRIS

2022

ACKNOWLEDGE

First of all, I want to sincerely thank my supervisor, Pn. Rohaizah binti Abdul Wahid, for her insightful advice and inspiring leadership during the course of this Final Year Project. I've gained a lot from your advice and knowledge. I also want to express my gratitude to Pn. Asma Hanee binti Ariffin and Encik Ahmad Nurzid bin Rosli, the lecturers, for their advice and assistance.

Also, I am very grateful to my responses for sharing their thoughts, suggestions, and encouragements to help the project's quality. I also like to thank my fellow friends for their help and recommendations during this assignment.

Last but not least, I would want to express my gratitude to my family, for always supporting me by giving encouraging words for me to finish my Final Year Project.

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

ABSTRACT

The accurate estimation of students' grades in future courses is important as it can inform the selection of next semester courses and create personalized degree pathways to facilitate successful and timely graduation. At present, students' dropout rate in university is gradually increasing and in the majority of cases drives the students to be either motivated or demotivated. Therefore MySPPS was developed to predict students' performance based on eight student's skills. For this purpose, Random Forest Regression had been used for classifying students' different levels of results and predicting students' performances. The result shows that RFR can perform with more than 80% accuracy. Thus, MySPPS provides decision-making support for students to choose courses reasonably to improve grades and remind them to understand their performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi
05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

CHAPTER 1

INTRODUCTION

1.1 Overview

In this era of advanced technology and the Internet, people's always rely on the technology to find solution for a problem and even get a suggestion on what they need to do in their life. Same goes to the important need of predicting student's performance that will benefit the students and the institute. Now a days, many university students struggle to finish their studies because they do not have proper plan on what subject they need to focus more in every semester in order for them to get a better result. It has been found that most of the students' drop-out from the universities during their first year is due to lack of proper support in under graduate courses (Callender and Feldman, 2009), (MacDonald, 1992). Due to this reason, the first year of the undergraduate student is referred as a "make or break" year. Without getting any support on the course domain and its complexity, it may demotivate a student and can be the cause to withdraw the course. There is a great need to develop an appropriate solution to assist students retention at higher education institutions. Early grade prediction is one of the solutions that have a tendency to monitor students' progress in the degree courses at the University and will lead to improving the students' learning process based on predicted grades.

To overcome this drawback, we suggest to develop a prediction system that can predict student's performance based on student's level of skills. Predictive modeling is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data. It is a crucial component of predictive analytics, a type of data analytic which uses current and historical data to forecast activity, behavior and trends. There are many choice of Machine Learning algorithm that we can use. Based on the research we have done, the most suitable algorithm that we can apply on the project is Random Forest Regression. Random forest is one of the best-performing learning algorithms. For social scientists, such developments in algorithms are useful only to the extent that they can access an implementation of the algorithm. Random forest has been shown to be a consistent high-performer in machine learning applications (Caruana & NiculescuMizil, 2006; Caruana, Karampatziakis, & Yessenalina, 2008; Fernandez-Delgado, Cernadas, Barro, & Amorim, 2014). Hardman, Paucar-Caceres, and Fielding (2013) applies random forest to identify inputs that best predict student progress from a large amount of student information system records. As what the past researcher have done and identify the suitable algorithm for the prediction, we come to the conclusion where the best algorithm that we applied on the project is Random Forest Regression.

1.2 Research Background

Machine Learning with EDM has gained much more attention in the last few years. Many machine learning techniques, such as Regression model (He, L. et al., 2018) are being used to predict students' grades. In this section, we will describe these machine learning techniques and how they are being used to predict students' grade in registered courses within the context of education. Random forest (Breiman, 2001) is an ensemble learning algorithm aimed at improving prediction accuracy through a forest of decisions trees. Random forest has been shown to be a consistent high-performer in machine learning applications (Caruana & Niculescu Mizil, 2006), Caruana et al., 2008) & (Fernandez-Delgado et al., 2014). However, random forest has seen very few applications in institutional research prediction tasks. Hardman, Paucar-Caceres & Fielding (2013) applies random forest to identify inputs that best

predict student progress from a large amount of student information system records. Langan et al. (2016) describes an approach using random forest to select bench marking factors to predict completion rates in nursing courses. The authors state that the utility of the method is appropriate for many forms of data at multiple scales. None of these previous studies focus discussion on the useful attributes of the random forest method other than prediction.

1.3 Problem Statement

Based on the study by Zafar Iqbal et al.(2017), most of the students' drop-out from the universities during their first year is due to lack of proper support in under graduate courses. This problem occur when the students did not understand which subject that they are easy to score and which subject that they need to focus more based on their skills. For example, if the students skill are more to communication skills and lack of programming skills, but did not focus more on the programming subject, at the end, it will effect the student programming subject result, because they do not know what to focus on unless there are system that help them to predict their performance and from there, they know what subject they need to focus more.

Based on the past research, when choosing pre-register courses students are usually uncertain because they do not know which ones are most suitable for them (Tran et al., 2017). Choosing the right course is a critical decision and it is important to get it right, as it can impact students' future success. Students enrolled in a course they are not happy with, typically study it with low motivation. That is why the important need of the system that can help the students to choose the right courses so that they are more motivated and more eager to learn without any worry about the subject that they are taken. Also, based on the predicted results, we can provide them early feedback, thus, we can prevent the dropping rate (or even expelling) every year.

The other problem that we can identify is that, there is no existing feature in MyUPSI Student system to predict students performance based on students skills. This is one of the major reason why we decided to build the system so that, we can

suggest to the university to add a new feature on the system specifically on the advisor function. Currently on MyUPSI Student system, on advisor function, the advisor can only view the students performance. If the advisor are from the same faculty and have the knowledge about what courses their students learn, it would be easier for the advisor to identify what are their students strength and weakness. But, it would be a problem if the advisor are from another faculty and do not have any knowledge on the field of study. The advisor cannot advise their student on what subject need to give more attention during the semester. With the existent of the system, the advisor can advise the student what subject that they can take on the next semester, and for the current semester, the advisor can advise the student what subject that they need to give more attention without the need of the advisor to have the knowledge on the field of study.

1.4 Research Objectives

1. To identify the the students academic performance based on their skills.

2. To develop MySPPS : The Development of Student's Performance Prediction System Using Machine Learning.

3. To test the functionalities of the developed system (MySPPS : The Development of Student's Performance Prediction System Using Machine Learning.)

1.5 Scope of Project

The project scope of this research is to predict students academic performance based on their skills type. The user scope of this study is Universiti Pendidikan Sultan Idris (UPSI) students degree in Software Engineering. The data scope of this study is UPSI academic result for students degree in Software Engineering.

1.6 Significant of Project

This research aim to help university students especially UPSI students to know their predicted result for that current semester so that they can change their learning style and focus on subject that are predicted in a low grade. It would also enable instructors to identify such individuals who might need assistance in the subject that need more attention. By having this kind of system, the students can have a better academic result by the end of the semester and it will increase the university academic performance.

1.7 Conclusion

In conclusion, the existing of prediction system is very important in education field. We can say that every university in Malaysia must have this system because it really help the students to change their learning style in every subject and also can help students advisor to know which subject they can advise for the students to focus and the advisor can inform the lecturer that teach the subject to give more attention to the students that potentially will get a lower result or failed the subject. Thus, it is necessary to build a system to predict student's academic performance.

In the following chapter, we will review and explain previous research on related topics and methodology to conduct the whole project. In chapter 5, we will analyse the research findings from the developed prediction system. In chapter 6, conclusion and future work will be discussed.