SOME PATTERN RECOGNITIONS FOR A RECOMMENDATION FRAMEWORK FOR HIGHER EDUCATION STUDENTS' GENERIC COMPETENCE DEVELOPMENT USING MACHINE LEARNING

Joseph Chi-ho So[1] ID **, Adam Ka-lok Wong**[1] ID **, Kia Ho-yin Tsang**[1] ID **, Ada Pui-ling Chan**[2] ID **, Simon Chi-wang Wong**[1] ID **, Henry C.B. Chan**[3] ID

[1]School of Professional Education and Executive Development,
The Hong Kong Polytechnic University (Hong Kong)
[2]Hong Kong Community College, The Hong Kong Polytechnic University (Hong Kong)
[3]Department of Computing, The Hong Kong Polytechnic University (Hong Kong)

*joseph.so@cpce-polyu.edu.hk, adam.wong@cpce-polyu.edu.hk, kia.tsang@speed-polyu.edu.hk,
ada.chan@cpce-polyu.edu.hk, simon.wong@cpce-polyu.edu.hk, henry.chan.comp@polyu.edu.hk*

Abstract

The project presented in this paper aims to formulate a recommendation framework that consolidates the higher education students' particulars such as their academic background, current study and student activity records, their attended higher education institution's expectations of graduate attributes and self-assessment of their own generic competencies. The gap between the higher education students' generic competence development and their current statuses such as their academic performance and their student activity involvement was incorporated into the framework to come up with a recommendation for the student activities that lead to their generic competence development. For the formulation of the recommendation framework, the data mining tool Orange with some programming in Python and machine learning models was applied on 14,556 students' activity and academic records in the case higher education institution to find out three major types of patterns between the students' participation of the student activities and (1) their academic performance change, (2) their programmes of studies, and (3) their English results in the public examination. These findings are also discussed in this paper.

Keywords – Classification and clustering, Supervised, unsupervised learning.

To cite this article:

----

1. Introduction

Generic competence (GC) is widely accepted as a critical element in the development of students in higher education (HE). Awareness of the importance of GC for both degree and sub-degree students continues to grow (Tait & Godfrey, 1999; Barrie, 2007; Duggan, 2014; So, Lam & So, 2013).

Student activities, including co-curricular and extra-curricular activities, which are collectively called GC development activities (GDAs), are important elements of GC development (Chan, 2010; Nghia, 2017). Unlike the formal curriculum, engagement in activities that support whole-person development is usually much less structured. The diversity of activities and students' freedom of choice are much greater than those in the formal curriculum. Moreover, the participation in GDAs is also dependent on the students' personalities. For example, Shiah et al. (Shiah, Huang, Chang, Chang & Yeh, 2013) showed that students' personalities can affect their involvement in extra-curricular activities and the development of their career skills.

HE students have to plan for their campus life and engage in the types of activities that will help them to realise their lifetime goals. Otherwise, they may miss promising opportunities or waste time and effort on participating in the activities that do not match their developmental needs, leading to the adverse effects on their academic studies (Torenbeek et al., 2010). To facilitate the HE students' planning for their campus life and GDAs, it is better for the students to understand the status of their GC, the student activities available as training and learning opportunities and the amount of time and effort required.

The HE students can choose to participate in a wide variety of GDAs. However, how do they know what activities are suitable for them? Not only may the students themselves not have a clear idea, but their advisors may not either. As there are many dimensions and factors to consider, the advice given to a student relies heavily on the advisor's personal experience, so subjectivity is inevitable. It is overwhelming to further provide a tailor-made GC development plan for each student as it involves great effort to gather information on each student's background and needs. Hence, there are many operational difficulties in providing sound recommendations to students. In this regard, a systematic evidence-based framework would greatly facilitate the whole operation.

In the commercial environment, data analytics (DA) and machine learning (ML) techniques are widely used in recommendation systems to improve the suitability of products or services recommended for a potential customer. These commercial applications reveal the potential of DA and ML for improving different areas of the education environment, and their application in HE has aroused great interest in academia (Popenici & Kerr, 2017). The advancement of such technologies can support GC development in a way that meets the needs of both the individual and the institution. Many DA and ML technologies are proposed to be used in HE. However, the relevant technology to facilitate GC development in tertiary institutions has been limited to simple provision, promotion and management of learning activities. To support and facilitate the development of GC, the challenges encountered by stakeholders, including students, advisors, activity organizers and providers, and institutions, need to be overcome. Platforms with a well-developed framework for applying learning support mechanisms, which include identifying the gap between students' GC development needs and their current statuses and consolidating all important information on individual students to support their selection of activities, could help to address these issues. However, such a platform cannot achieve this goal without a clear framework that supports students' GC development and that can consider both their aspirations and institutional values. Obviously, a gap exists in HE between the application of technology and the framework needed to support GC education.

2. Research Objective and Significance

To address the deficiencies in the literature, this project will aim to formulate a recommendation framework in terms of development activities that lead to students' GC. Furthermore, despite the success of the recommendation system (RS) mechanism in commercial online platforms, the model cannot be directly used in HE. In a commercial context, success is measured by the number of user subscriptions rather than the benefits for end-users. However, in education, students' actual needs and whole-person development are the key. A unique recommendation is needed for each student, so the gap between GC development and the students' current situation should be incorporated in the framework.

By bringing ML technologies to a new paradigm, a framework for a recommendation system to help tertiary students develop their GC to achieve their career goals will be developed. This study will be

beneficial for students, especially freshmen, as well as lecturers, professional advisors and educational institutions. Students' development of GC in HE will be reinforced by their newly acquired ability to select the activities most suitable for their personal development, particularly in the area of GC.

3. Methodology

In order to make the recommendation system work as intended, pattern recognition will be an important part of the proposed recommendation system. Pattern Recognition (PR) is an intelligent behaviour that is based on patterns to describe or classify certain measurements (Schalkoff, 2007). Using PR helps the research to narrow down the features for finding out the principal components of different target features, thus improving the efficiency and feasibility of the recommendation system.

4. Pattern Recognition in Machine Learning–Enabled Recommender Systems

Machine learning (ML) is a branch of artificial intelligence whose goal is to find meaningful patterns in a large data set using computational models without being explicitly programmed to perform the task. An ML application will apply a computation model to a large set of data, adjust the parameters automatically and find the best possible pattern for predictive purposes. ML has been used by businesses such as Amazon and YouTube, whose systems automatically recommend books and videos based on the data from all previous visitors to the website. Recommendation systems are also widely used in online platforms for marketing and promotion purposes (Tarus, Niu & Yousif, 2017).

PR is shown to be important and feasible in a machine learning enabled system. As stated by Bulgarevich et al. (Bulgarevich, Tsukamoto, Kasuya, Demura & Watanabe, 2018), by applying PR with ML, a system that is efficient, accurate and quick to recognize the patterns of metallurgical microstructures in optical microscopy images can be produced.

5. Pattern Recognition for GDA Involvement

An understanding of students' behaviour is critical to predict and address future students' involvement in learning activities. To understand how the students' behaviour addresses their involvement in learning activities, 14,556 students' activity and academic records in the case HE institution were collected and analysed. DA and PR mechanisms in the data mining tool Orange with some programming in Python were applied to reveal patterns of student involvement in various activities, especially those oriented towards GC improvement and academic achievement. We seek to discover whether the students tend to choose activities that enhance the GCs at which they already excel, those at which they need to improve, or those required to meet the academic programme's needs. DA techniques, including K-means and Bayesian methods, incorporated into Orange were used to detect clusters, group data and carry out association rule learning to discover the relationships among the variables of the student activities in the data sets. These techniques have been used in other contexts, including identification of students who may fail in some academic subjects (Akçapınar, Altun & Aşkar, 2019). The correlations between various factors, the development of specific GC and the types of activities were also identified via factor analysis and correlation analysis.

The data sources included, but not limited to the following:

- Demographics - gender and age.

- Academic background - subjects and grades in high school (e.g. Hong Kong Diploma of Secondary Education Examination (HKDSE)).

- Current studies – Academic programs and subjects taken, including discipline-specific and general education subjects at the institution of the investigators.

- Co-curricular Achievement Records – Records of participation of GCA for all students

- Activity nature: theme, duration, intended learning outcomes (ILOs), modes – The details of all activities are available from the Student Affairs Office, and it is noted that each activity is associated with one or two ILOs.

- Institution's expectations of graduate attributes – In this study, Hong Kong Community College of the Hong Kong Polytechnic University (PolyU) is used as the case study. The learning outcomes of sub-degree degree level of PolyU are Competent Paraprofessional, Critical Thinker, Effective Communicator, Practical Problem Solver, Lifelong Learner and Ethical Citizen (Learning Outcomes for PolyU Graduates of HD Programmes, 2018).

- Graduate Survey - The experiences of graduates in previous cohorts.

- A self-assessment questionnaire on GC - feedback from the self-assessment on all-around development (SAARD)

6. Tests

Tests were performed with the aid of Orange with some programming in Python to find out possible principal components of ML models. The aim of the tests is to discover the correlations and patterns among the features and target features, thus minimizing the resources and time needed to train the ML model and maximizing efficiency when training and using the model.

7. Analytical Results

Tests were carried out on the collected students' activity and academic records to find out the patterns between features and target features using the machine learning models, namely, linear regression and polynomial regression. These machine learning models were programmed in Python and integrated with Orange. The results reveal the following three major patterns:

7.1. Finding 1 - Patterns between Participation of Activities and Cumulative GPA Change

This test attempted to find out the pattern between the students' participation of the student activities and the change in cumulative GPAs. The cumulative GPA change is calculated by finding the difference between two semesters, regardless of whether the difference is positive or negative. In this test, the difference in the cumulative GPA is calculated with each student's cumulative GPAs of semester 1 and semester 2 in 2018. The participation of student activities is the total number of non-one-off activities a student took part in. Table 1 presents the terms used and their corresponding meanings. Figure 1 and 2 present the patterns by linear regression and polynomial regression respectively.

Terminology	Meaning
Y1 = CumlGPAChange.2018.1-2	Change in cumulative GPA between semester 1 and semester 2 in 2018
Y2 = CumlGPAChange.201819.2-1	Change in cumulative GPA between semester 2 in 2018 and semester 1 in 2019
Y3 = CumlGPAChange.2019.1-2	Change in cumulative GPA between semester 1 and semester 2 in 2019
Z1 = CumlGPAChange.2018.1-2(Cont.)	Change in cumulative GPA between semester 1 and semester 2 in 2018, continuous data
Z2 = CumlGPAChange.201819.2-1(Cont.)	Change in cumulative GPA between semester 2 in 2018 and semester 1 in 2019, continuous data
Z3 = CumlGPAChange.2019.1-2(Cont.)	Change in cumulative GPA between semester 1 and semester 2 in 2019, continuous data
x1 = ('One-off Count', 'Non-One-off Total')	Total participation number of Non-One-off activities
x2 = ('One-off Count', 'One-off Total')	Total participation number of One-off activities

Table 1. Terminology of Graphs in Finding 1

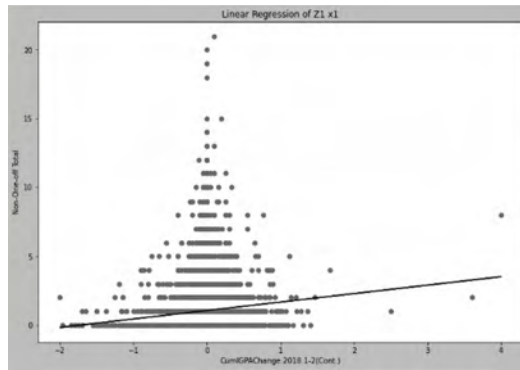### 7.1.1. Linear Regression Results

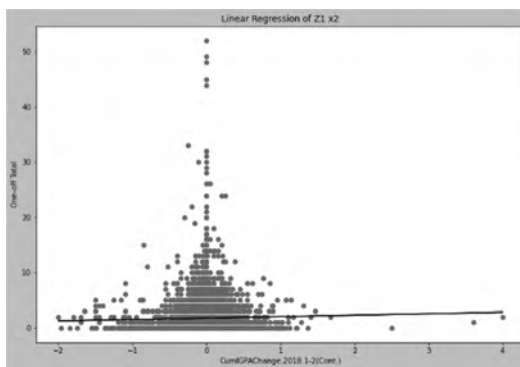Figure 1. Linear Regression of Z1 x1 in Finding 1

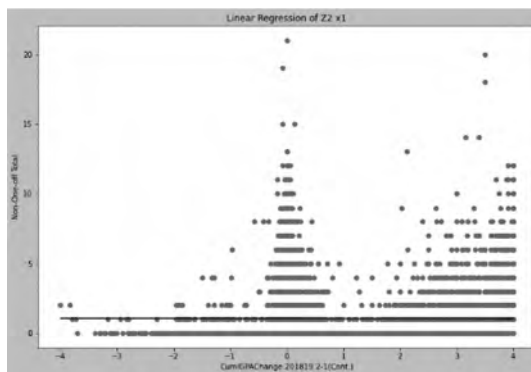Figure 2. Linear Regression of Z1 x2 in Finding 1

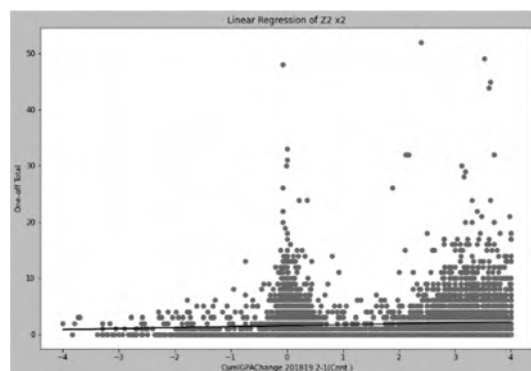Figure 3. Linear Regression of Z2 x1 in Finding 1
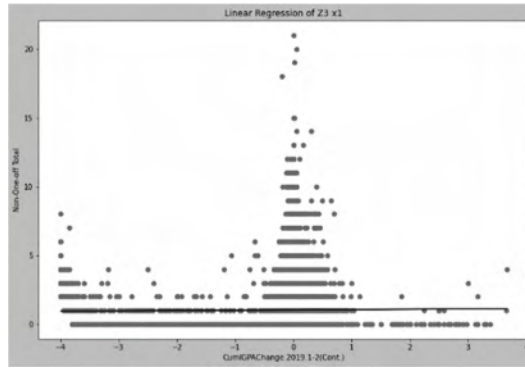

Figure 4. Linear Regression of Z2 x2 in Finding 1

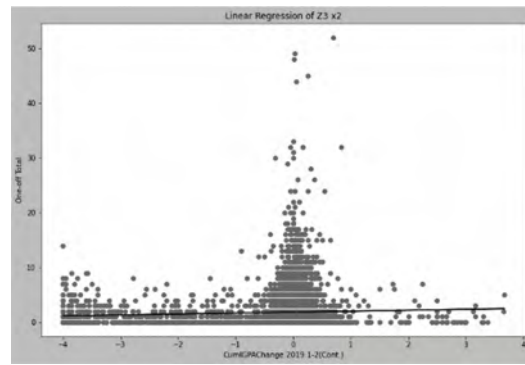Figure 5. Linear Regression of Z3 x1 in Finding 1

Figure 6. Linear Regression of Z3 x2 in Finding 1
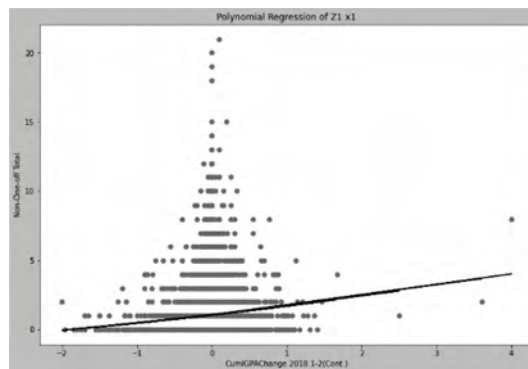
7.1.2. Polynomial Regression Results
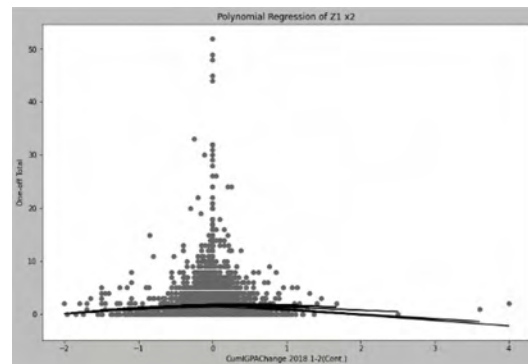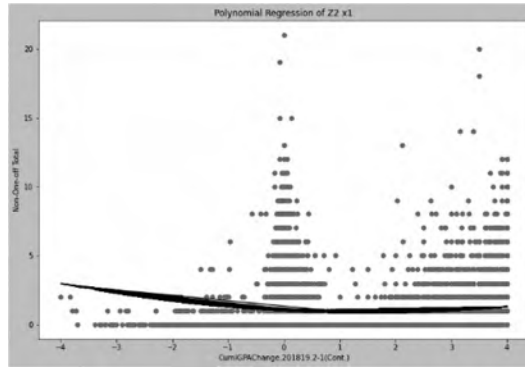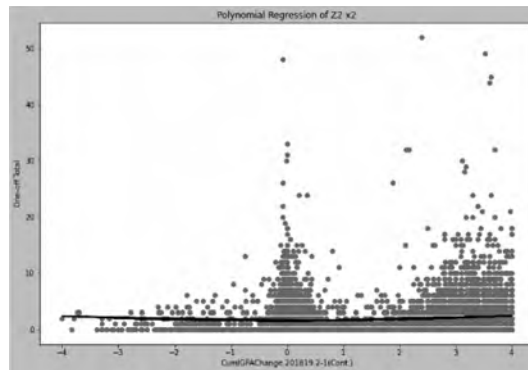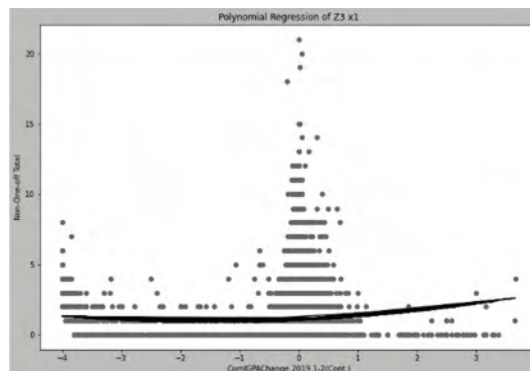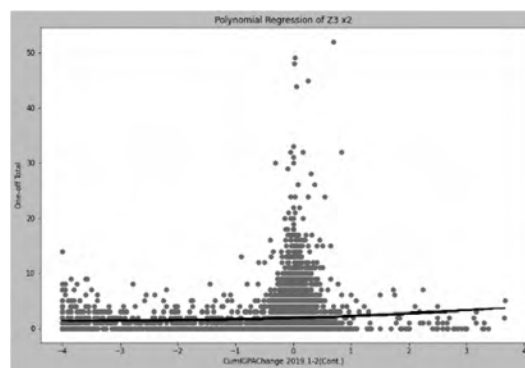

Figure 7. Polynomial Regression of Z1 x1 in Finding 1

Figure 8. Polynomial Regression of Z1 x2 in Finding 1

Figure 9. Polynomial Regression of Z2 x1 in Finding 1

Figure 10. Polynomial Regression of Z2 x2 in Finding 1

Figure 11. Polynomial Regression of Z3 x1 in Finding 1

Figure 12. Polynomial Regression of Z3 x2 in Finding 1

From Figures 1 to 12, the graphs showed a small trend that when the total participation of Non-One-off activities is higher, the change in cumulative GPA from semester 1 to semester 2 is also higher.

7.2. Finding 2 - The Patterns of the Participation of Activities with Programme of Each Student

The programmes were encoded for Python to analyse with Linear Regression and Polynomial Regression. In Figures 3 and 4, X in the graphs represents the programmes; y1 represents the feature "('One-off Count', 'Non-One-off Total')"; and y2 represents the feature "('One-off Count', 'One-off Total')". Table 2 shows the terms used and their corresponding meanings.

Terminology	Meaning
X = Programme	Programme that the student is studying
y1 = ('One-off Count', 'Non-One-off Total')	Total participation number of Non-One-off activities
y2 = ('One-off Count', 'One-off Total')	Total participation number of One-off activities

Table 2. Terminology of Graphs in Finding 2
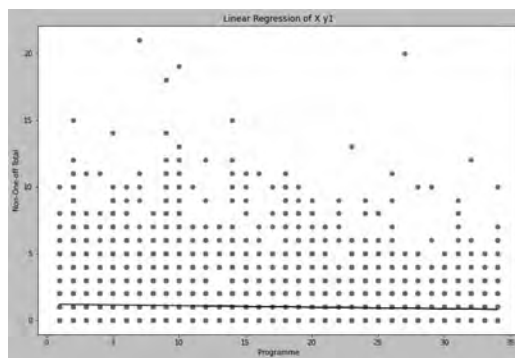
7.2.1. Linear Regression Results

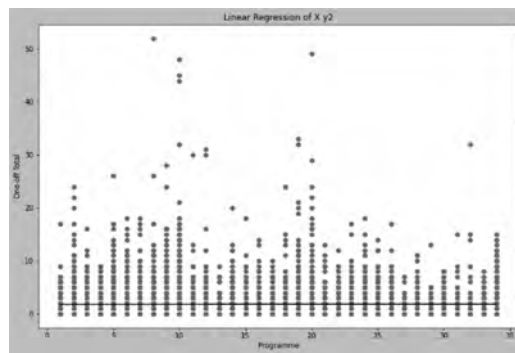Figure 13. Linear Regression of X y1 in Finding 2

Figure 14. Linear Regression of X y2 in Finding 2

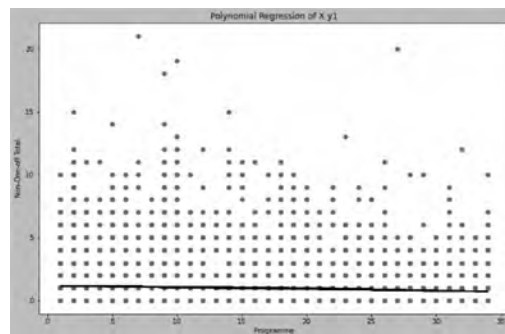**7.2.2. Polynomial Regression Results**

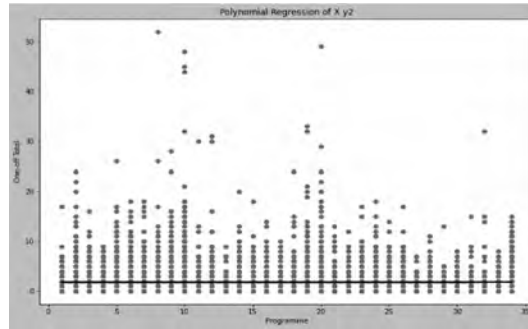Figure 15. Polynomial Regression of X y1 in Finding 2

Figure 16. Polynomial Regression of X y2 in Finding 2

From the Figures 13 to 16, the regression lines show no trend between programmes, One-off and Non-One-off activities.

7.3. Finding 3 - Patterns between Participation of Activities and the Results of the Subject English in Public Exam

The results of the subject English in the Hong Kong Diploma of Secondary Education (HKDSE) are used to find patterns with the participation of activities. Table 3 shows the terms and their corresponding meanings used in Figures 5 and 6.

Terminology	Meaning
X = ENG	The results of the subject English in the Hong Kong Diploma of Secondary Education (HKDSE)
y1 = ('One-off Count', 'Non-One-off Total')	Total participation number of Non-One-off activities
y2 = ('One-off Count', 'One-off Total')	Total participation number of One-off activities

Table 3. Terminology of Graphs in Finding 3

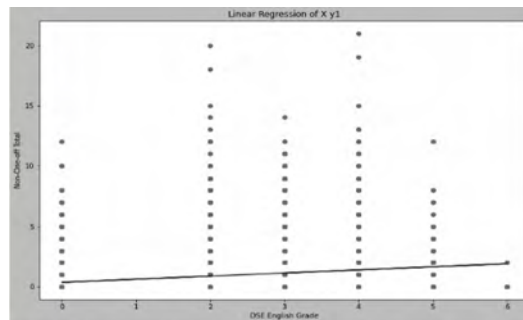### 7.3.1. Linear Regression Results

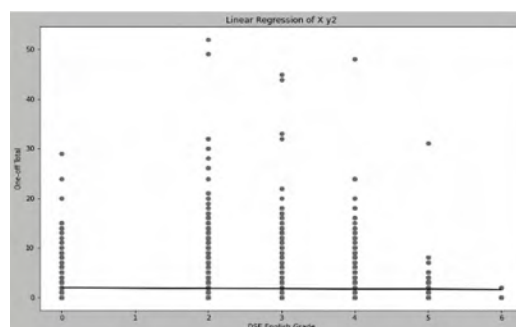Figure 17. Linear Regression of X y1 in Finding 3

Figure 18. Linear Regression of X y2 in Finding 3

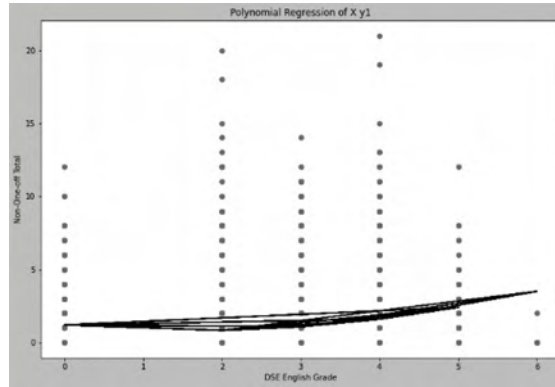### 7.3.2. Polynomial Regression Results

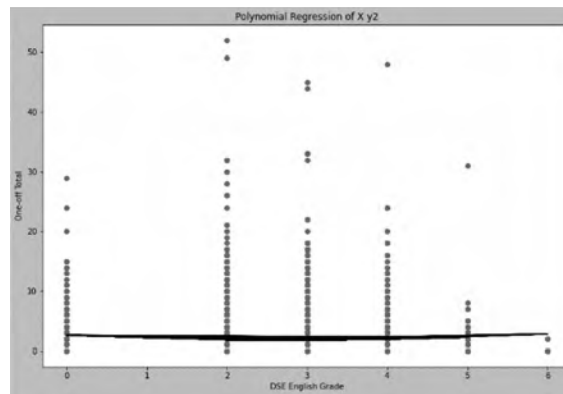Figure 19. Polynomial Regression of X y1 in Finding 3

Figure 20. Polynomial Regression of X y2 in Finding 3

From Figures 17 to 20, a small trend can be seen that students will participate more Non-One-off activities when they achieve higher grade in the HKDSE English subject. However, the trend is not significant. There is no significant difference between higher and lower grades in the subject English when interpreting the graph of the participation of One-off activities and the grade of the subject English.

8. Discussion

In finding 1, the graphs showed that there are no huge impact on the changes in cumulative GPA that are related to the participation of both One-off and Non-One-off activities, meaning that the participation of activities should not affect student's GPA.

In finding 2, the flat lines indicates that no students in specific programmes have a higher chance to participate in any activities.

In finding 3, the trend found is not significant. There is no significant difference between higher and lower grades in the subject English when interpreting the graph of the participation of One-off activities and the grade of the subject English.

9. Conclusion

In this paper, we aim to find the patterns between the students' participation of the student activities and their academic performance change, their programs of studies, and their English results in the public examination. No significant patterns was found thus indicating that students' academic performance change, programmes of studies, and their English results in the public examination have no significant

effect on the participation activities. The findings can help eliminating the features that are not as significant when training the ML model.

Declaration of Conflicting Interests

Funding

References

Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1). https://doi.org/10.1186/s41239-019-0172-z

Barrie, S.C. (2007). A conceptual framework for the teaching and learning of generic graduate attributes. *Studies in Higher Education*, 32(4), 439-458. https://doi.org/10.1080/03075070701476100

Bulgarevich, D.S., Tsukamoto, S., Kasuya, T., Demura, M., & Watanabe, M. (2018). Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures. *Scientific Reports*, 8(1). https://doi.org/10.1038/s41598-018-20438-6

Chan, W.S.C. (2010). Students' understanding of generic skills development in a university in Hong Kong. *Procedia, Social and Behavioral Sciences*, 2(2), 4815-4819. https://doi.org/10.1016/j.sbspro.2010.03.776

Duggan, L. (2014). A quantitative analysis of students' perception of generic skills within an undergraduate electronics/mechanical engineering curriculum. *Online Submission*. https://eric.ed.gov/?id=ED546781

Learning Outcomes for PolyU Graduates of HD Programmes (2018). PolyU.Edu.Hk. Available at: https://www.polyu.edu.hk/obe/institutional_policies/Learning_Outcomes_for_PolyU_Graduates_of_High_Diploma_Programmes.pdf

Nghia, T.L.H. (2017). Developing generic skills for students via extra-curricular activities in Vietnamese universities: Practices and influential factors. *Journal of Teaching and Learning for Graduate Employability*, 8(1), 22-39. https://doi.org/10.21153/jtlge2017vol8no1art624

Popenici, S.A.D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22. https://doi.org/10.1186/s41039-017-0062-8

Schalkoff, R.J. (2007). Pattern Recognition. In *Wiley Encyclopedia of Computer Science and Engineering*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470050118.ecse302

Shiah, Y.J., Huang, Y., Chang, F., Chang, C.F., & Yeh, L.C. (2013). School-based extracurricular activities, personality, self-concept, and college career development skills in Chinese society. *Educational Psychology*, 33(2), 135-154. https://doi.org/10.1080/01443410.2012.747240

So, J.C.H., Lam, S.Y., & So, Y.L. (2013). A case study of generic competencies among science and technology tertiary graduates in Hong Kong. *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. https://doi.org/10.1109/TALE.2013.6654484

Tait, H., & Godfrey, H. (1999). Defining and assessing competence in generic skills. *Quality in Higher Education*, 5(3), 245-253. https://doi.org/10.1080/1353832990050306

Tarus, J.K., Niu, Z., & Yousif, A. (2017). A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generations Computer Systems: FGCS*, 72, 37-48. https://doi.org/10.1016/j.future.2017.02.049

Torenbeek, M., Jansen, E., & Hofman, A. (2010). The effect of the fit between secondary and university education on first-year student achievement. *Studies in Higher Education*, 35(6), 659-675. https://doi.org/10.1080/03075070903222625