RGC Ref. No.:
UGC/FDS24/E09/20
(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR
THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines:
1. *Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
2. *Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1. Project Title

A Machine Learning Model for Recommendation System for Generic Competency Development in Higher Education

2. Investigator(s) and Academic Department(s) / Unit(s) Involved

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	Dr WONG, Adam Ka-lok / Senior Lecturer	Division of Science, Engineering and Health Studies / PolyU SPEED
Co-Investigator(1)	Ir Dr SO, Chi-ho / Chief Technology Officer	Smart Location Technology Limited
Co-Investigator(2)	Dr CHAN, Chun-bun / Associate Professor & Associate Head	Department of Computing / PolyU
Co-Investigator(3)	Dr CHAN, Pui-ling / Director of Marketing And Business Development	Smart Location Technology Limited
Co-Investigator(4)	Dr WONG, Simon Chi-wang / Lecturer	Division of Science, Engineering and Health Studies / PolyU SPEED
Others	Mr TSANG, Ho-yin / Research Assistant	General Office / PolyU SPEED

3. Project Duration

	Original	**Revised**	**Date of RGC / Institution Approval** *(must be quoted)*
Project Start Date	01/01/2021	N/A	N/A
Project Completion Date	31/12/2023	N/A	N/A
Duration *(in month)*	36	N/A	N/A
Deadline for Submission of Completion Report	31/12/2024	N/A	N/A

4.3 Please attach photo(s) of acknowledgement of RGC-funded facilities / equipment.
N/A

Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. To learn and discover the pattern of students' involvement in GDAs by applying DA and pattern recognition techniques using the data from the previous cohorts.

2. To discover the impact on the student participation behaviour in GDAs with other factors about the students (including personality, career aspiration, gap of GC needs and learning experiences in high school) in the coming cohorts in addition to the factors found in Obj1.

3. Incorporate the findings from above Obj1 and Obj2 to develop an RS framework in GDAs using an ontological approach.

4. To evaluate the effectiveness of various recommendation algorithms, including collaborative filtering and content-based recommendation, in providing systematic recommendations on the selection of GDAs.

5. To evaluate the effectiveness of the recommendation framework by assessing user acceptance and the improvement in GC of a group of selected students via various assessments.

5.2 Revised objectives

Date of approval from the RGC: N/A

Reasons for the change:

1. Not applicable

2.

3. ....

5.3 Realisation of the objectives

(Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

Objective 1: Fully Achieved

To learn and discover the pattern of students' involvement in GDAs by applying DA and pattern recognition techniques using the data from the previous cohorts.

We analyzed 14,018 records of students' participation in GDAs . The records covered three cohorts, and we discovered that the pattern of students' involvement in GDAs could be represented as a sparse matrix after applying feature engineering to quantity the duration of student participation.

Objective 2: Fully Achieved

To discover the impact on the student participation behaviour in GDAs with other factors about the students (including personality, career aspiration, gap of GC needs and learning experiences in high school) in the coming cohorts in addition to the factors found in Objective 1.

To the dataset from objective 1, we added students' HKDSE results, academic performance at PolyU HKCC, self-assessment of their GC levels, personality and career aspirations in terms of John Holland's Six Types of Personality. Then we used various machine learning algorithms, plus genetic algorithms, and ensemble methods such as stacking to analyse the impact of these factors on students' participation in GDAs.

Objective 3: Fully Achieved

Incorporate the findings from above Obj1 and Obj2 to develop an RS framework in GDAs using an ontological approach.

The dataset from objective 2 was about 98 features from 9570 students. We applied the Principal Component Analysis and SelectKBest algorithms to reduce the number of features to 8. We found that the reduced-feature dataset retained similar accuracy for prediction while saving processing time. This dataset was used for the next objective.

Objective 4: Fully Achieved

To evaluate the effectiveness of various recommendation algorithms, including collaborative filtering and content-based recommendation, in providing systematic recommendations on the selection of GDAs.

Using the dataset from objective 3, we compared the performance of the collaborative filtering and content-based algorithms. We found that LightFM, a hybrid model, performed best. We built a recommendation system (RS) using LightFM.

Objective 5: Fully Achieved

To evaluate the effectiveness of the recommendation framework by assessing user acceptance and the improvement in GC of a group of selected students via various assessments.

A group of students used the RS during their orientation sessions. Their acceptance was measured by a survey. A group of students were invited to a teamwork exercise and a quiz on GC which was developed for this project. Their performance was collated with their GDA participation to assess their improvements.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. To learn and discover the pattern of students' involvement in GDAs by applying DA and pattern recognition techniques using the data from the previous cohorts.	✓	100%
2. To discover the impact on the student participation behaviour in GDAs with other factors about the students (including personality, career aspiration, gap of GC needs and learning experiences in high school) in the coming cohorts in addition to the factors found in Obj1.	✓	100%
3. Incorporate the findings from above Obj1 and Obj2 to develop an RS framework in GDAs using an ontological approach.	✓	100%
4. To evaluate the effectiveness of various recommendation algorithms, including collaborative filtering and content-based recommendation, in providing systematic recommendations on the selection of GDAs.	✓	100%
5. To evaluate the effectiveness of the recommendation framework by assessing user acceptance and the improvement in GC of a group of selected students via various assessments.	✓	100%

6. Research Outcome

6.1 Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

The term Generic Competency (GC) refers to skills and abilities essential to succeed in one's work across multiple disciplines. General Community Development Activities (GDAs) are essential elements of GC development. The research started by investigating the pattern of students' involvement in General Community Development Activities (GDA) using data from students in a community college in Hong Kong. After data pre-processing and feature engineering, there were 14,018 valid records records with 111 features. The students' involvement was measured in terms of the number of GDAs participating. We employed statistical methods, data analytics, and machine learning to perform methodological triangulation. In particular, we used ANOVA, various machine learning algorithms such as ANN, KNN, and random forest, genetic algorithms, and ensemble machine learning methods such as stacking to analyse the data. We found that students with strong secondary school academic backgrounds are more engaged in GDAs. We found that the student involvement in GDAs was very low for most students and that their academic performance at the college was not negatively affected by their degree of involvement in the GDAs. It was also found that the programme of study had no association with the students' involvement in GDAs. It proved that the GDAs were suitable for all students in various programmes.

We found that when proper features were selected, the number of features for machine learning could be reduced to shorten data processing time and enhance result interpretation. We collected 98 features from 9570 students and applied the Principal Component Analysis and SelectKBest algorithms to reduce the number of features to 8. Finally, we compared the accuracy of predictions using KNN and ANN based on the all-feature dataset with those based on the reduced-feature dataset. The results showed that the reduced feature dataset maintained good prediction accuracy for building recommendation systems to recommend the GCDAs to students.

Based on the reduced features, we developed MARS, which stands for a machine learning-assisted recommendation system. The machine learning model of the MARS was trained using school advisors' recommendations of GDAs for more than 400 student profiles in two previous cohorts. The collaborative filtering and content-based recommendation algorithms were compared when training the model. After comparing various machine learning algorithms such as KNN, Random Forests, and so on, we found that LightFM, a hybrid model, provided the best performance. The MARS was given to new incoming students to evaluate its effectiveness in making recommendations. The student's acceptance of the MARS was examined using a survey using the constructs from the models of Task-Technology Fit (TTF), Unified Theory of Acceptance and Use of Technology (UTAUT) and Hierarchical Model of Leisure Constraints (HMLC). We found that the students had positive perceptions about the MARS. Finally, we compared the improvements in the GCs among current students with two previous cohorts using different assessment methods, both self-assessment and quiz assessment. It was found that there is an optimum number of GDAs for students to improve their GCs. Students who joined 1 to 5 GDAs had higher GCs than students who did not join any GDA and students who joined six or more GDAs.

Our findings were published in four peer-reviewed journal articles and eight international conference papers. Among the four journals, which one journal ranked Q1 on the Scimago Journal & Country Rank, while two ranked Q2. The details of these papers can be found in Part C of this report.

6.2 Potential for further development of the research and the proposed course of action *(Maximum half a page)*

The focus groups revealed that the student's choices for General Community Development Activities (GDAs) are influenced by a variety of dynamic and personal factors. These factors include, but are not limited to, their financial status, class schedules, and peer influence. However, the research used historical data from student admission surveys which did not explicitly ask these factors. Furthermore, it was noted that some students may not join an activity that they were very interested in but did not join due to the factors stated above. Unfortunately, these factors were not included as features in the training data. We suggest that future research should incorporate these factors as features to provide more impactful results. Furthermore, we suggest to conduct a longitudinal study in tracking the perceptions and ~~GC~~ improvements in the generic competencies of students who followed the MARS recommendations.

The recommendation system can be further developed to be integrated with the institution's activity administration and promotion system. At present, all students in the institution receive the same promotional materials for GDAs. When the suggested integration is completed, each student can receive a personalisation promotion message that includes a rating from MARS about the fit of the GDA to the personality of the student.

7. Layman's Summary

(Describe <u>in layman's language</u> the nature, significance and value of the research project, in no more than 200 words)

The term generic competency (GC) refers to skills and abilities essential to succeed in one's work across multiple disciplines. GC includes teamwork, problem-solving, communication, and critical thinking skills. GC development activities (GDAs) are essential elements of GC development. However, participation in GDAs is voluntary and much less structured than a formal curriculum. Students in higher education have limited time to take up GDAs. Therefore, students in higher education need advisors to help them choose GDAs that match their personalities and career aspirations. This research developed a machine learning-assisted recommendation system (MARS) for GDAs in higher education. The machine learning model of the MARS was trained using school advisors' recommendations of GDAs for more than 400 student profiles in two previous cohorts. Then, the MARS was given to new incoming students to evaluate its effectiveness in making recommendations. The survey and focus group results revealed that the students had positive perceptions about the MARS. A group of students provided self-assessed GC ratings and joined a quiz that assessed their GC. The results showed that students who took part in GDAs that matched the recommendation from MARS showed improvements in their GCs.

Part C: Research Output

8. Peer-Reviewed Journal Publication(s) Arising Directly From This Research Project

(Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				**Author(s)** *(denote the corresponding author with an asterisk*)*	**Title and Journal / Book** *(with the volume, pages and other necessary publishing details specified)*	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowl-edged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
Year of Publication	**Year of Acceptance** *(For paper accepted but not yet published)*	**Under Review**	**Under Preparation** *(optional)*						
2023	N/A	N/A	N/A	*Joseph Chi-Ho So, Yik Him Ho , Adam Ka-Lok Wong , Henry C. B. Chan , Kia Ho-Yin Tsang, Ada Pui-Ling Chan, and Simon Chi-Wang Wong	Analytic Study for Predictor Development on Student Participation in Generic Competence Development Activities Based on Academic Performance, IEEE Transactions On Learning Technologies, VOL. 16, NO. 5, OCTOBER 2023 https://doi.org/10.1109/TLT.2023.3291310	Yes / 2022	Yes (Appendix 1)	Yes	Yes
2023	N/A	N/A	N/A	Adam Ka Lok Wong*, Joseph Chi Ho So, Kia Ho Yin Tsang, Ran Wei	Selecting the Best K Features for Predicting Student Participation in Generic Competency Development Activities in Higher Education, International Journal of Emerging Technologies in Learning Vol. 18 No. 23 (2023), https://doi.org/10.3991/ijet.v18i23.45499	No	Yes (Appendix 2)	Yes	Yes

The Latest Status of Publications				Author(s) *(denote the corresponding author with an asterisk*)*	Title and Journal / Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowl-edged the Support of RGC *(Yes or No)*	Accessible from the Institutional Repository *(Yes or No)*
Year of Publication	**Year of Acceptance** *(For paper accepted but not yet published)*	**Under Review**	**Under Preparation** *(optional)*						
2023	N/A	N/A	N/A	*Joseph Chi-Ho So, Adam Ka-Lok Wong, Kia Ho-Yin Tsang, Ada Pui-Ling Chan, Simon Chi-Wang Wong, Henry C. B. Chan	Some Pattern Recognitions for a Recommendation Framework for Higher Education Students' Generic Competence Eevelopment using Machine Learning. Journal of Technology and Science Education, 13(1), 104-115. https://doi.org/10.3926/jotse.1707	No	Yes (Appendix 3)	Yes	Yes
2024	N/A	N/A	N/A	*Simon Chi-Wang Wong, Adam Ka-Lok Wong, Yui-Yip Lau, Kia Ho-Yin Tsang, Ada Pui-Ling Chan	Perceived usefulness of a machine learning-assisted recommendation system for generic competency development. Journal of Education and e-Learning Research, 11(3), 614-621. https://doi.org/10.20448/jeelr.v11i3.5971	No	Yes (Appendix 4)	Yes	Yes

9. Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered

(Please attach a copy of each conference abstract)

Month / Year / Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowl-edged the Support of RGC *(Yes or No)*	Accessible from the Institutional Repository *(Yes or No)*
Dec 2021/HK	Data Analytic Framework on Student Participation in Generic Competence Development Activities	IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) 2021	Yes / 2022	Yes (Appendix 5)	Yes	Yes

Month / Year / Place	**Title**	**Conference Name**	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowl-edged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
Dec 2021/HK	Machine Learning Assisted Planning in Student Development Activities Involvement	International Conference on Learning and Teaching 2021 (ICLT2021)	Yes / 2022	Yes (Appendix 6)	Yes	Yes
Jan 2022/HK	Feature Engineering for a Building a Machine Learning Model for Recommendation System for Generic Competency Development in Higher Education	CPCE Centre for Pedagogic Research Conference 2022	Yes / 2022	Yes (Appendix 7)	Yes	Yes
Jul 2023/ Toronto	Personality Types of RIASEC Theory Among Students' Choice of Disciplines in Education	Canada International Conference on Education (CICE 2023)	No	Yes (Appendix 8)	Yes	Yes
Jun 2023/ Holland	Impact of the Six Personality Types Among RIASEC Theory on the Student Participation of Extra-Curricular Activities	International Academic Conference on Teaching, Learning and Education (TLECONF 2023)	No	Yes (Appendix 9)	Yes	Yes
Jun 2023/ Holland	SelectKBest and PCA To Pre-process The Dataset to Enhance Prediction in the Student Participation in Extracurricular Activities	International Academic Conference on Teaching, Learning and Education (TLECONF 2023)	No	Yes (Appendix 10)	Yes	No
Dec 2023 / Singapore	The Development & Evaluation of a Machine Learning-Assisted Recommendation System for Generic Competencies Development	2023 International Conference on Big Data Analysis and Application	No	Yes (Appendix 11)	Yes	No
Apr 2024 / London	Evaluating the Effectiveness of a Machine Learning Model for Recommendation System for Generic Competency Development in Higher Education	2024 EdTec – International Conference on Education & Learning Technology, 22-23 April, London	No	Yes (Appendix 12)	Yes	Yes

10. Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning
(Please elaborate)

This research gained new knowledge in using advanced machine learning algorithms such as collaborative filtering, matrix factorisation and LightFM. These algorithms, and their applications in recommendation systems like the ones in this research, contributed to enriching the teaching of the courses SEHS4696 Machine Learning and Data Mining, and SEHS4678 Artificial Intelligence.

11. Student(s) Trained
(Please attach a copy of the title page of the thesis)

Name	Degree Registered for	Date of Registration	Date of Thesis Submission / Graduation
	BSc (Hons) in Applied Sciences (Information Systems & Web Technologies)	1 September 2020	29 April 2022
	Bachelor of Science (Honours) Scheme in Applied Sciences	1 September 2021	28 April 2023
	Bachelor of Science (Honours) Scheme in Applied Sciences	1 September 2021	28 April 2023

12. Other Impact
(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

Best Presentation Award at 2023 International Conference on Big Data Analysis and Application

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (please specify)	
No. of outputs arising directly from this research project	4	8	N/A	N/A	Type N/A	No.

14. Public Access Of Completion Report

(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	**Reasons**
N/A	