Towards the Development of a Career Path Recommender System for Senior High School in Selected Public Schools using Multi-Label Classification

Abelardo T. Bucad

University of Makati

ABSTRACT

The Philippines challenged accepting reforms that shifted 10-year to the k-12 basic education program, adding two years to the students' basic education. Each graduate student from the grade 10 level must choose a career path that will give them the skills they need to be future-ready. Given the long-term effects on a student's career, selecting the SHS track or strand is one of the most important life decisions they will ever make. In this paper, a career path recommender system will introduce to avoid a mismatch of strands/tracks and help the students choose the possible career path in Senior High School (SHS) that fits their taste and capability. The researcher will collect related data from six (6) selected public schools offering grade 10 in Valenzuela City. And will explore different student data such as family background, big five personalities, grades, IQ test, and some tools in the Multiple Development Toolkit from Department of Education (Self-assessment and career interest cluster, Personal career exploration). The study will use content-based filtering and classification algorithm in machine learning like Decision trees (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Support Vector Machine (SVM) that will serve as base-classifier in the multi-label classification.

Index Terms: About four keywords or phrases in alphabetical order, separated by commas.

INTRODUCTION

In 2016, the Philippines challenged accepting reforms that shifted 10-year to the k-12 basic education program. The new curriculum includes Senior High School (SHS) or grades 11 and 12. The focus of the SHS curriculum is to prepare the students for further studies, employment, entrepreneurship, and middle-skills development. SHS consists of four specialization tracks: the academic track, technical-vocational-livelihood track, arts and design, and sports track. Some of these tracks are divided into different strands. Selecting the SHS strand is one of the learner's most significant challenges because of its long-term impact on their career. It is associated with socioeconomic, quality of life, and social status [1]. Therefore, having a good career choice at the early stage can achieve a better future in the chosen career and better economic growth. On the other hand, the wrong career choice could affect the fate of an individual, financial status, waste precious time, and increase investment costs for learners' education.

K to 12 programs allowed graduates to become competitive in their chosen field of specialization and possess the necessary skills and competencies [2]. However, there is a noticeable situation in which students enroll in college courses that are not associated with their track or strand of senior high school. The mismatch rate is increased between the strand of the learners during their senior high school and the course they enrolled in college. It is revealed that for the school year 2017-2018, the mismatch rate is thirty-nine percent (39%) and thirty-one percent (31%) for the SY 2018-2019)[3]. This track or strand of senior high school mismatch could lead to job mismatch in the future. A recent study by the Philippine Institute for Development Studies (PIDS) revealed that 39% of employed Filipinos are "overeducated." At the same time, a quarter is "undereducated," highlighting the seriousness of the country's education-job mismatches

[4]. That is why President Ferdinand Marcos Jr.'s administration reviews education career guidance programs to check what areas need to be reinforced to address job mismatch [5].

Data mining (DM) is a statistical process that uses mathematical and machine learning techniques and algorithms to extract information from databases [6]. Recently, there has been a lot of interest in using data mining techniques in education. It has been a powerful tool for finding hidden patterns in educational data like student information, academic history, skills, and exam outcomes. These data become helpful to the data scientist in turning them into meaningful information to provide predictions and recommendations. DM and Recommendation system can make significant changes in the current situation in secondary education in the Philippines in terms of assessment and proper counseling in selecting appropriate careers for students to increase the competitiveness of the student graduates.

PURPOSE OF THE STUDY

Since there are still problems with choosing the track and the k 12 program is relatively new in the Philippines. The study aims to solve social issues while also elevating the study by applying operationalizing system theories to help students choose the right path in the future. A career path recommender system will introduce to avoid mismatch of strands/tracks and help the students choose the possible career path in SHS that fits their taste and capability.

PRELIMINARY LITERATURE REVIEW

Different recommender systems were developed related to education and students' career. Shu [7] proposed a content-based recommendation algorithm based on a convolutional neural network (CNN). The CNN is being used to predict the latent factors from the text information of the multimedia resources.

Amer [8] explain how to design and evaluate a recommender system using KNN and Naïve Bayes Classification algorithm. The papers follow a content-based approach by building user profile knowledge and action and then comparing it to course attributes to recommend courses. In 2019, Kurniadi [9] proposed a framework for an intelligent recommender system for college students. It is a data mining and machine learning technique for predicting student performance and providing a subject recommendation. It uses clustering techniques, association rules, and classification using Support Vector Machine (SVM), Naïve Bayes, and k-Nearest Neighbor(k-NN) in analyzing students' academic records. The study reveals that machine learning and data mining can be recommender systems.

The work of Natividad [10] proposed a career recommender system for senior high school students in K to 12 education. This aim is to help the students decide on what career to pursue. It uses different filter methods to select attributes for CRISP inputs. The Family background, Final grades., Parents/peer/relative influence, and socio-economics status as the dataset of 459 students from the Philippines.

The fuzzy logic model produced an RMSE of 4.82 in the testing set. This model would help the students in career decisions. Still, some factors need to be considered, such as feature selection techniques, using other algorithms and collecting more data to generate good results. Another fuzzy-based algorithm Qamhieh [11] developed in Palestine is PCRS: Personalized Career-Path Recommender System for Engineering Students.

 This help the high school students choose engineering discipline using fuzzy intelligence of N-layered architecture using students' academic performance, personality type, and extra-curricular skills. The dataset collected 1250 engineering students from different engineering disciplines. Academic performance, personality type, and extra-curricular skills are essential to generating a personalized recommendation system. However, the PCRS is only capable of students interested in pursuing engineering courses. Verma [11] developed a Student career path recommendation in the engineering stream based on a three-dimensional model by employing fuzzy logic and influence in skill questions for a career path. Hence the model plays a significant role in generating the desired output, but this can only predict for engineering

courses. Another system developed in Nigeria using machine learning is the Autonomous course recommender system. It aims to reduce the wrong placement of applicants by recommending appropriate courses for students based on Senior High school grades. It collected data from 8700 students from two universities that employed five machine learning classification models, including Linear Regression, Naïve Bayes, SVM, KNN, and Decision trees algorithm. This study has made good prediction models but needs to consider other students' attributes in recommending the course [13].

The work of Izadi [14] uses Naïve Bayes and Logistic regressions classifier to predict software topics from GitHub using a multi-label classification method. The dataset includes textual information such as descriptions, README files, wiki pages, and file names. Jain [15] developed a system for predicting consumer travel and tourist recommendations. A multi-label classification approach is used to create the recommendation system. The classification method was implemented using K-Nearest Neighbors, Support Vector Machine, Multilayer Perceptron, Logistic Regression, Random Forest, and Ensemble Learning to train the model.

Several recommender systems have been explored in education, just like the work of Shu [16] who collected 10393 books as a dataset to recommend learning resources to the students in China. The work of Morsomme [17] proposed a course recommender system for the Liberal Arts bachelor's program at the University College Maastricht in the Netherlands using content-based. It aims to provide academic counseling and assist students in selecting courses with more excellent knowledge. The system recommends courses based on the content best matches the student's academic interests, course description, and past academic performance (grades) in terms of Kullback Leibler distance. Other studies like Mokarrrama [18] provide a recommendation system to help prospective students choose the most suitable private universities for admission. The dataset consists of GPA, the secondary school certificate examination, university tuition, ratings, and ranking evaluated with the 947 feedback from the students and obtained an acceptable accuracy using f1 score, precision, recall, and balance accuracy performance metrics. Content-Based Filtering techniques (CBF) have been subjected to various machine learning approaches.

Recently, many machine learning-related software-based algorithms have been proposed using a content-based filtering approach [19]. Amer [20] developed a course recommender system using KNN and Naïve Bayes. Tanya [21] proposed Career Recommendation Systems using Content-based Filtering using the K-means algorithm to help college graduates fulfill their dreams by recommending a job based on their interests and skillset.
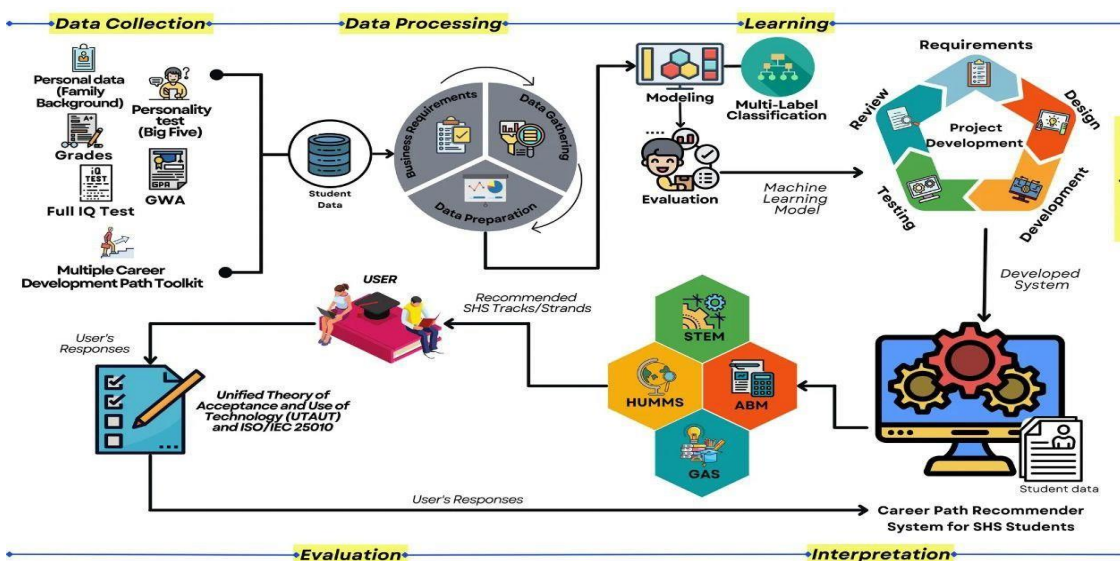
Conceptual Framework

Fig. 1: Conceptual Model of the Study [22]

Figure 1 shows the diagram of a Conceptual Framework of the study. The conceptual model represents the overall point of the research. It depicts the flow of data, and its processing before the users use the system. It shows that the researcher will collect all necessary student data from selected public schools, including student data(family background), personality trait results, grades, some assessment tools from MCDP, and IQ test results. After gathering, all data will be processed through CRISP-DM, including business requirements, data gathering, preparation, modeling, model evaluation, and deployment. In the part of the modeling phase, a Multi-label classification will apply. The next process will follow the software development life cycle to develop the system, including planning, designing, developing, testing, and releasing.

The development of the system will use relevant programming languages and methods that comply with the system's design. When it comes to testing, the process will start with creating and testing through created test cases. Then the final product will be evaluated using ISO/IEC 205010 for software and UTAUT for user behavioral intention on the system.

RESEARCH QUESTIONS

A career path is a series of jobs leading to your short- and long-term career goals. Some career paths are linear paths within one field. In contrast, others change domains periodically to attain career or personal goals either through the industry or your track through the organization. The researcher intends to mine the student's career path/track and develop a recommender system that helps the students reach their career goals. The researcher sought the following:

1. What significant attributes contribute to the students predicting the future Senior High Career strands/track?

Independent Variables:

1.1 Student data (Family Background)

1.2 Personality test (Big Five Personality Trait)

1.3 General Weighted Average (average of all subjects)

1.4 Grades(per subject)

1.5 Multiple Career Development Path Toolkit (Self-assessment and career interest cluster)

1.6 Full IQ Test

Dependent Variable:

1. Multiple Career Development Path Toolkit (Personal career exploration)
2. What supervised machine learning algorithms are best suited in choosing the Senior High Career Strands/Track that fits students' capabilities?
3. How can data mining techniques be used to discover the patterns in choosing the Senior High Career Strands/Track?
4. How will the recommender application software be evaluated in terms of quality using the ISO/IEC 25010:2011 and be evaluated regarding user intentions on the use of technology using the Unified Theory of Acceptance and Use of Technology (UTAUT)?

OBJECTIVE OF THE STUDY

The study's main objectives are to design, develop, and test the career path recommender system for senior high schools in selected public schools to help guidance counselors, parents, and students choose the right path in the future. The following are the specific objectives:

1. To identify the attributes that significantly contribute to the students to predict the future Senior High Career strands/track:

Independent Variables:

- Student data (Family Background)
- Personality test (Big Five Personality trait)
- General Weighted Average (Average of all subjects)
- Grades (per subject)
- Multiple Career Development Path Toolkit (Self-assessment, career interest cluster)
- Full IQ Test

Dependent Variables:

1. Multiple Career Development Path Toolkit (Personal career exploration)
2. To determine the supervised machine learning best suited in choosing the Senior High Career Strands/Track that fits students' capabilities.
3. To determine the discover patterns in choosing the Senior High Career Strands/Track using applicable data mining technique.
4. To evaluate the recommender application software in terms of quality using the ISO/IEC 25010:2011 and the user intention on the use of technology using the Unified Theory of Acceptance and Use of Technology (UTAUT).

PROPOSED METHODOLOGY

This chapter aims to introduce the research methodology for this quantitative theory and whether the researcher has grounds to say that data mining techniques and multi-label classification can be used in this study. This part explains on the different methods to prove the techniques' efficiency. This chapter states the researcher's approach in terms of research and development of the study.

Research Design

The researcher will use a descriptive quantitative method and developmental research. Descriptive research aims to obtain accurate and detailed information about the characteristics, trends, and behavior of the phenomena [22]. Descriptive quantitative research collects quantifiable data from the population sample for statistical analysis. It will use a quantitative approach to explore new knowledge of the student's career path/track in the Senior High school dataset from selected public schools in Valenzuela City. Developmental research is a product development process that collects, analyzes, develops, and evaluates developed programs. In several scholarly works by [23], [24], and [25] used developmental research.

Research Procedure

The researcher will use classification data mining techniques to find a significant pattern from the student's data. It also employs Cross-Industry Standard Process for Data Mining (CRISP-DM) to meet the study's goal.

The process of models includes Business understanding, data understanding, data preparation, modeling,

evaluation, and deployment phase. At the Business Understanding stage, the researcher will set the primary goal of this study and should answer what to apply/develop in the system, which is to provide an application that will help the student recommend a career path that matches their strengths and skills. The researcher will explore career-related factors, machine learning, and other techniques to meet the goal. At the Data Understanding stage, the researcher will collect data from the public students composed of several attributes such as family background, grades, IQ, Personality tests, and multiple career development path results. These data will analyze in the python programming interface using different statistical tools to gain insight into the dataset. The Data Preparation stage covers activities to improve the dataset from initial raw data to the final dataset. At this phase, the researcher will provide a dataset of independent and dependent variables based on the multi-label classification requirements. The researcher will perform data processing techniques such as data cleaning, feature selection, and data transformation to generate the model in the next stage of this process. In the Modeling phase, the researcher will use a content-based filtering method and various multi-label classification techniques such as binary relevance, label power set, and classifier chain. Each method will apply Naive Bayes, Trees, Support Vector Machines, K nearest-neighborhood, and Trees as a based-classifier to build the model. At the evaluation stage, the researcher will assess the model results to meet the business objectives in the evaluation stage. To determine the appropriate models, the researcher will evaluate the accuracy of the algorithms based on the Accuracy Score, Hamming loss, Precision and Recall, and F1 Score. After the evaluation trained model is successful, the model integrates into the Deployment phase. In this stage, the researcher will summarize the deployment strategy, including the necessary steps to perform the software development methodology stated in this study.

Respondents of the Study

The respondents of the study of the six (6) selected public schools in Valenzuela City for software quality are six (6) principals, six (6) program coordinators, six (6) guidance counselors, and six (6) technical experts. And 10% of the student's estimated population will be the respondents of the user intentions on the use of information systems using the proportionate stratified random sampling with a total of fifty-four (254) respondents.

Data Gathering and Statistical Analysis

This study needs to collect data from the student, which includes personal and sensitive information. To do this, the researcher will write a letter to the Department of Education in Valenzuela to allow the researcher to collect this information. The letter covers the collection of grade 10 students with assurance to follow the data privacy act of 2012 in processing data. The researcher will collect data thru secondary data. ISO/IEC 25010:2011 adapted from Gatpandan [26] without modification will be used as a research instrument to evaluate the recommender system for system quality. The researcher will use frequency count and weighted mean in the descriptive analysis. The tally or frequency count calculates how many people fit into a specific category or the number of times a characteristic occurs. This calculation is expressed by both the absolute (actual number) and relative (percentage) totals. The weighted mean is calculated by multiplying the weight (or probability) associated with a particular event or outcome with its quantitative outcome and then combining all the products.

The UTAUT adapted from Venkatesh [27] would be the additional research instrument that will be used in this study. In this instrument, the Partial Least Squares-Structural Equation Modeling (PLS-SEM) method using Smart PLS software will be employed to determine the student's behavioral intention toward the system. This method is the most popular research tool for data analysis for non-normal data, small sample sizes, and the use of formative indicators [28].

CONCLUSION

Based on the related studies stated in this paper, numerous studies have been conducted on the career path recommender system. These systems provide opportunities and solutions to various career problems. These

can play an important role for students as a guide in selecting a future career. These systems collected data ranging from 456 to 8700 from students based on the student performance and skill interest as the dataset for the model. These systems explore machine learning and data mining and use content-based filtering in their studies. The major machine learning techniques of classification and statistics are commonly used across these studies using different algorithms such as Decision tree, KNN, SVM, and Naïve Bayes as based-classifier for the recommender system.

While reviewing related studies, there are some gaps have been identified:

1. The researcher often focuses only on factors such as student academic performance and skills interest in the studies that used machine learning techniques.
2. Although theories on career development are already well-established, very little research has examined how these theories relate to people's early career track decisions. Factors influencing the choice of career path are still poorly understood, particularly in the Philippine setting.
3. They only use multi-class classification in producing output.

The researcher is inspired to develop a recommender system that recommends senior high school tracks and strands. The proposal will use a content-based filtering and classification algorithm in machine learning like SVM, Naïve Bayes, KNN, and Decision trees to serve as base classifiers in multi-label classification. In addition, aside from academic performance and career interest, the study will explore additional factors for recommending systems such as family background, Big five personalities, IQ test results, and some Multiple career development tool kits from the Department of Education as a dataset for the classification model.

REFERENCES

1. J. J. G. Ouano, J. F. dela L. Torre, W. I. Japitan, and J. C. Moneva, "FACTORS INFLUENCING ON GRADE 12 STUDENTS CHOSEN COURSES IN JAGOBIAO NATIONAL HIGH SCHOOL – SENIOR HIGH SCHOOL DEPARTMENT," International Journal of Scientific and Research Publications (IJSRP), vol. 9, no. 1, p. p8555, Jan. 2019, doi: 10.29322/ijsrp.9.01.2019.p8555.
2. R. L. Dizon, J. S. Calbi, J. S. Cuyos, and M. Miranda, "Perspectives on the Implementation of the K to 12 Program in the Philippines: A Research Review," International Journal of Innovation and Research in Educational Sciences, vol. 6, no. 6, pp. 2349–5219, 2019.
3. C. Quintos, D. G. Caballes, and M. R. Valdez, "Perceptions of Teachers on the Different Strains of Online Modality of Learning The Digital Skills of Secondary School Teachers in Manila View project Assessment of Students in Digital Game-Based Learning in Teaching Physics 7 View project." [Online]. Available: https://www.researchgate.net/publication/349176836
4. M. v Tutor, A. C. Orbeta, and J. M. Miraflor, "The 4th Philippine Graduate Tracer Study: Examining Higher Education As Pathway To Employment, Citizenship, and Life Satisfaction from the Learner's Perspective." [Online]. Available: https://www.pids.gov.ph
5. MakarlikaTv, "UniTeam to address education-job mismatch among Pinoy workers," https://maharlika.tv/2022/03/04/uniteam-to-address-education-job-mismatch-among-pinoy-workers/.
6. A. J. P. Delima, A. M. Sison, and R. P. Medina, "Variable reduction-based prediction through modified Genetic Algorithm," International Journal of Advanced Computer Science and Applications, vol. 10, no. 5, pp. 356–363, 2019, doi: 10.14569/ijacsa.2019.0100544.
7. J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A content-based recommendation algorithm for learning resources," Multimed Syst, vol. 24, no. 2, pp. 163–173, Mar. 2018, doi: 10.1007/s00530-017-0539-8.
8. Amer S. El-Ameer and Ammar A. Neamah, Design and Evaluation of a Course Recommender System Using Content-Based Approach. 2018.
9. D. Kurniadi, E. Abdurachman, H. L. H. S. Warnars, and W. Suparta, "A proposed framework in an intelligent recommender system for the college student," in Journal of Physics: Conference Series,

Dec. 2019, vol. 1402, no. 6. doi: 10.1088/1742-6596/1402/6/066100.

10. M. C. B. Natividad, B. D. Gerardo, and R. P. Medina, "A fuzzy-based career recommender system for senior high school students in K to 12 education," in IOP Conference Series: Materials Science and Engineering, Mar. 2019, vol. 482, no. 1. doi: 10.1088/1757-899X/482/1/012025.

11. M. Qamhieh, H. Sammaneh, and M. N. Demaidi, "PCRS: Personalized Career-Path Recommender System for Engineering Students," IEEE Access, vol. 8, pp. 214039–214049, 2020, doi: 10.1109/ACCESS.2020.3040338.

12. P. Verma, S. K. Sood, and S. Kalra, "Student career path recommendation in engineering stream based on three-dimensional model," Computer Applications in Engineering Education, vol. 25, no. 4, pp. 578–593, Jul. 2017, doi: 10.1002/cae.21822.

13. M. Isma'Il, U. Haruna, G. Aliyu, I. Abdulmumin, and S. Adamu, "An Autonomous Courses Recommender System for Undergraduate Using Machine Learning Techniques," Mar. 2020. doi: 10.1109/ICMCECS47690.2020.240882.

14. M. Izadi, A. Heydarnoori, and G. Gousios, "Topic recommendation for software repositories using multi-label classification algorithms," Empir Softw Eng, vol. 26, no. 5, Sep. 2021, doi: 10.1007/s10664-021-09976-2.

15. P. K. Jain, R. Pamula, and E. A. Yekun, "A multi-label ensemble predicting model to service recommendation from social media contents," J Supercomput, vol. 78, no. 4, pp. 5203–5220, 2022, doi: 10.1007/s11227-021-04087-7.

16. J. Shu, X. Shen, H. Liu, B. Yi, and Z. Zhang, "A content-based recommendation algorithm for learning resources," Multimed Syst, vol. 24, no. 2, pp. 163–173, Mar. 2018, doi: 10.1007/s00530-017-0539-8.

17. R. Morsomme and S. Vazquez Alferez, "Content-based Course Recommender System for Liberal Arts Education."

18. M. J. MOKARRAMA, S. KHATUN, and M. S. AREFIN, "A content-based recommender system for choosing universities," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 28, no. 4, pp. 2128–2142, Jul. 2020, doi: 10.3906/ELK-1911-37.

19. J. and S. S. S. Khan Aslam Hasan and Siddqui, "A Survey of Recommender Systems Based on Semi-supervised Learning," in International Conference on Innovative Computing and Communications, 2022, pp. 319–327.

20. Amer S. El-Ameer and Ammar A.Neamah, Design and Evaluation of a Course Recommender System Using Content-BasedApproach. 2018.

21. Tanya V Yadalam, Vaishnavi M Gowda, Vanditha Shiva Kumar, and Disha Girish, Career Recommendation Systems using Content based Filtering.

22. Canva.com., Free Design Tool: Presentations, Video, Social Media | Canva," https://www.canva.com/(accessed : Sept, 20,2022).

23. Shona McCombes., "What Is a Research Design | Types, Guide & Examples," https://www.scribbr.com/methodology/research-design/, 2022.

24. C. Jin et al., "Development and evaluation of an artificial intelligence system for COVID-19 diagnosis," Nat Commun, vol. 11, no. 1, p. 5088, 2020, doi: 10.1038/s41467-020-18685-1.

25. T. A. Arshi, S. Islam, and N. Gunupudi, "Predicting the effect of entrepreneurial stressors and resultant strain on entrepreneurial behaviour: an SEM-based machine-learning approach," International Journal of Entrepreneurial Behavior & Research, vol. 27, no. 7, pp. 1819–1848, Jan. 2021, doi: 10.1108/IJEBR-08-2020-0529.

26. N. Burton, M. Burton, C. Fisher, P. G. Peña, G. Rhodes, and L. Ewing, "Beyond Likert ratings: Improving the robustness of developmental research measurement using best-worst scaling", doi: 10.3758/s13428-021-01566-w/Published.

27. P. H. Gatpandan and S. C. Ambat, "Mining Disciplinary Records of Student Welface and Formation Office : An Exploratory Study to Enhance the University Services Portfolio," International Journal of Information Technology, Control and Automation, vol. 7, no. 2, pp. 23–38, Apr. 2017, doi: 10.5121/ijitca.2017.7202.

28. V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information

Technology: Toward a Unified View," MIS Quarterly, vol. 27, no. 3, pp. 425–478, May 2003, doi: 10.2307/30036540.

29. M. Lock, S. Yee, and M. S. Abdullah, "A Review of UTAUT and Extended Model as a Conceptual Framework in Education Research," Jurnal Pendidikan Sains Dan Matematik Malaysia, vol. 11, pp. 1– 20, 2021, doi: 10.37134/jpsmm.vol11.sp.1.2021.