



INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv



Students Demography Clustering Based on The ICFL Program Using K-Means Algorithm

Rachmadita Andreswari^{a,*}, Rokhman Fauzi^a, Berlian Maulidya Izzati^a, Vandha Pradwiyasma Widartha^a,
Dita Pramesti^a

^a Department of Information System, Telkom University, Bandung, 40257, Indonesia

Corresponding author: *andreswari@telkomuniversity.ac.id

Abstract— Independent Campus, Freedom to Learn (ICFL) Program is one of the manifestations of student-centered learning. This program can help students reach their full potential by allowing them to pursue their passions and talents. This study aims to see how the segmentation of students participating in the ICFL program is based on demographic data. This research is based on survey responses from students participating in the ICFL program. The method used in this study is input data preparation, pre-processing, data cleansing, and data analysis. The information will be pre-processed before being utilized and evaluated. To help produce better outcomes in data clustering, the K-Means clustering approach is used, which is processed using the Python computer language. The data is clustered using the K-Means clustering approach based on gender characteristics, Grade Point Average (GPA), university entrance selection, ICFL category, and year or semester when participating in ICFL. This study resulted in three clusters with each of its criteria. The dominant gender is found in clusters 2 (100% female) and 3 (100% male). Software Development was the most popular ICFL category among students in cluster 1, accounting for 67%, while Design and Analysis Information Systems was the most popular in clusters 2 and 3. The most dominant ICFL program is found in three clusters. ICFL - Internship program in which at least 40% of participants come from each cluster. The research results are expected to assist stakeholders in evaluating the implementation of the ICFL program.

Keywords— ICFL; independent campus freedom to learn; clustering; higher education; K-Means.

Manuscript received 21 Mar. 2022; revised 16 Sep. 2022; accepted 12 Dec. 2022. Date of publication 30 Jun. 2023.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

University employment makes up only 10.18 percent of total employment as of February 2021. One of the difficulties students encounter is the disruption of technology in the workplace, which has resulted in various job opportunities [1]. From these issues, colleges must develop an original learning approach that covers parts of attitudes, knowledge, and skills that are maximally and always applicable. Student-centered learning is one of the most important types of creative learning that may be used [2].

The Independent program Campus, Freedom to Learn (ICFL), which stands for Freedom of Learning - Independent Campus, was implemented by the Ministry of Education at the beginning of 2020. The ideology of two interrelated ideas, independence, and freedom, inspired Indonesian education's founder[3]. ICFL is a method of education used in an independent and adaptable institution to foster a culture of inventive learning that is driven by student needs rather than

imposing rules. Students have numerous opportunities through this curriculum to broaden and enhance their perspectives and real-world competencies through their beliefs and passions.

ICFL program is one of the Ministry of Education and Culture's programs to prepare students for social and cultural changes, the workplace, and rapid technological advancements. Because students are required to have competencies that align with the times' demand the program promotes student-centered learning [4]. The ICFL program is a manifestation of Minister of Education Regulation No. 3 of 2020, concerned with independent and flexible learning in higher education. The ICFL program also allows students to complete three semesters of study outside their program of study foster a creative, non-restrictive learning environment that meets the needs of students [4].

As a result, all academically-level universities, including Telkom University are actively modifying their curricula and enhancing the effectiveness of the SN-DIKTI learning

process while supporting the initiative. Considering the wide range of student residences and socioeconomic origins is important. So, to put the program into effect, a draft rule is required. (1) To give opportunities and fulfill the right of students to select relevant learning experiences based on their interests and their desire to master the learning objectives of one or more courses by involving students in lecture programs or extracurricular activities. (2) to develop knowledge on how to contextualize science in the workplace and/or community life, which enables students to gain additional value in enhancing and strengthening their skills or high-quality academic achievements. With the hopes that it will produce graduates equipped to handle the increasingly difficult problems of modern life [5].

Learning on a campus that operates independently offers difficulties and chances to build independence in learning through the realities and dynamics of the field. The students who have self-directed learning abilities will learn how to learn. Online learning requires the learner to be self-directed or to be able to guide and direct their learning [6]. The widespread consensus is that learners have greater control over their education when they learn online [7], [8]. One benefit of distance learning is that students can begin, stop, and modify their individual learning as necessary [9].



Fig. 1 Types of ICFL Programs [2]

Figure 1 shows the types of ICFL programs, which include internships/work practices in the industry or other workplaces, implementing community service projects in villages, teaching in educational units, participating in student exchanges, conducting research, conducting entrepreneurial activities, making independent studies/projects, and participate in humanitarian programs.

In a prior study, the School of Business, University of Nicosia, evaluated the Task-Based Internship program. Interns' interpersonal interactions and communication skills can be improved, and students can be better prepared for their future employment with this form of task-based, hands-on learning [10]. The experiences of 37 students who participated in industrial training are the topic of Ooi's research [11]. Students are encouraged to think of ways to

improve their experience. The purpose of this study was to evaluate how off-campus internships alter students' perceptions and self-evaluations, as well as the difference between expectations and reality [12], [13], [14].

Another study states that it is necessary to have an assessment and qualification document that is used as a reference/guideline that shows the level of mastery of competence by each student [15]. Other research suggested that UCH's faculty of science and engineering should provide more internship opportunities to their students [16]. It will be beneficial for students to become better professionals. According to Chickerur [17], internships with projects positively affect students in dimensions such as (i) increasing the likelihood of promotion and advancement in recruitment (ii) duration as a trainee after being hired (iii) benefits during the internship period, increasing confidence (iv) in the world work.

Research on the ICFL student exchange program during the Covid-19 pandemic revealed that student exchange management begins with planning, organizing, implementing, and evaluating [18]. This research can be used as a recommendation for other colleges to conduct online student exchanges. This study aims to discover how exchange students view change and development during student exchange within the context of identity [19].

Clustering is the process of gathering like objects together in one cluster and diverse objects together in another. Clustering's major goal is to use an unsupervised learning process to identify groups with similar characteristics and assign them to groups [20], [21]. The K-Means algorithm was used to cluster the data in this study. The K-Means algorithm is used because it is simple to construct and because an enhanced K-Means algorithm can aid in better student grouping results [22].

An examination of student internships majoring in preventive medicine at six colleges and universities in China was undertaken in a previous related study to provide reference suggestions for enhancing internship arrangements [23]. Another study looked at how data mining may be used to perform a simple grouping analysis to understand student behavior better. Data mining methods are a suitable benchmark for determining whether students' performance improves over time [24]. Clustering algorithms and various other statistical approaches are used to investigate the real outcomes of university exams for MCA (Master of Computer Application) students while examining student performance [25].

The K-Means algorithm is also used to group data in characterizing student impressions of project-based learning models in design courses and training program strategies [26]. Other studies have used the clustering method to classify engineering students based on their preferred learning styles [27]. Students seeking higher education in the Indian states of Punjab and Haryana were also analyzed using the clustering method [28].

This study focuses on the results of a survey completed by 83 students who participated in the ICFL program. This study uses the K-Means clustering method to see demographics based on gender, Grade Point Average (GPA), university admission selection path, ICFL category, and semester year when participating in ICFL. From the clustering, it is expected

to be able to show the segmentation of students participating in ICFL based on student demographic data.

II. MATERIAL AND METHOD

This study analyzes and recommends the applicability of the ICFL program using the CRISP-DM (Cross-Industry Standard Process for Data Mining) method. CRISP-DM is widely used in data mining and knowledge discovery projects to the point where it is considered a standard [29]. The CRISP-DM process model depicts the cycle's six steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [30].

CRISP-DM consists of six phases, and one is Business Understanding, which is a critical stage because it necessitates knowledge of the business, how to obtain data, and how to connect modeling goals with business goals. 2) Data Understanding is the process of looking at data to see if there are any problems. 3) Data Preparation involves cleaning or transforming data to address issues with it. 4) Modeling, which uses data mining tools and predetermined algorithms and processes to generate predictive or descriptive models. 5) Evaluation is performed to ensure that the previously generated model aligns with the objectives. 6) Deployment of the model or a plan to use it [30].

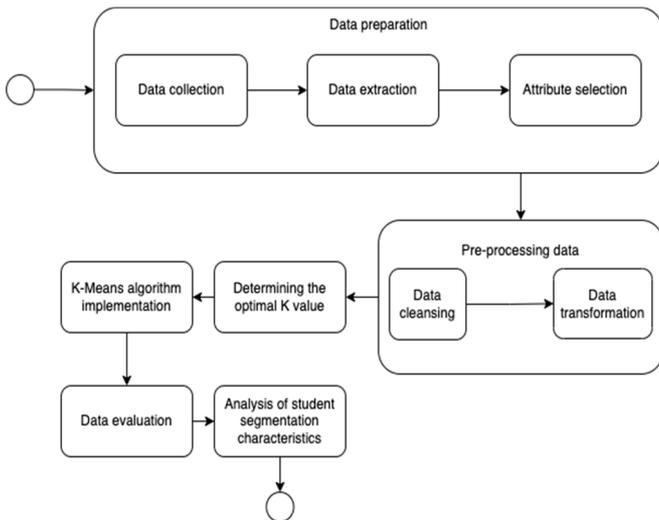


Fig. 2 Research method

The data for this study came from a survey of students who had taken part in the ICFL program. The Python programming language is used to handle data, and the K-Means algorithm is used because it can decrease variances within a cluster while maximizing variations with data from other clusters. The clustering findings will be assessed and examined in terms of the characteristics of the segmentation of students participating in the ICFL.

III. RESULTS AND DISCUSSION

This study relied on original data from an ICFL survey of 83 participants, with 53 percent of men and 47 percent of women. Figure 3 depicts the student distribution based on their participation in the ICFL program.

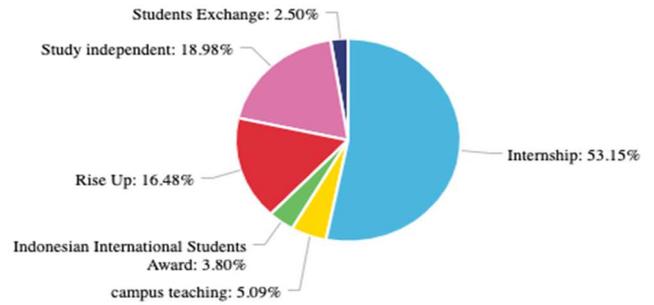


Fig. 3 Distribution of Students by ICFL Program

On survey data, exploratory data analysis (EDA) approaches were employed to determine the association between data attributes. One hot encoding was used to transform categorical data into numeric data, spectral embedding was used to determine the distribution of related features, and the K-Means algorithm was used to cluster or segment students participating in the ICFL program. GPA, gender, university admission selection path, ICFL category, and semester year engaging in ICFL activities are used to segment students.

A. One Hot Encoding

Data processing can be done correctly if the data being processed is of the same type and scale. Different data forms, especially for clustering or classification, will impact the predicted outcomes. One-Hot-Encode is a conversion method that turns a category variable into a new column with a 0 or 1 value (numeric variable)

Ethnic	Javanese	Sundanese
Javanese	1	0
Sundanese	0	1
Javanese	1	0
Sundanese	0	1
Javanese	1	0

Fig. 4 Implementing one hot encoding on data attributes

An example of implementing one hot encoding on data attributes of categorical type is shown in Figure 4

B. Spectral Embedding

Spectral embedding is used to group data attributes that have commonalities by understanding the affinity or similarity matrix or eigenvalues in data that has been transformed into the same data scale. Figure 5 shows how to organize data attributes based on how similar their dimensions are.

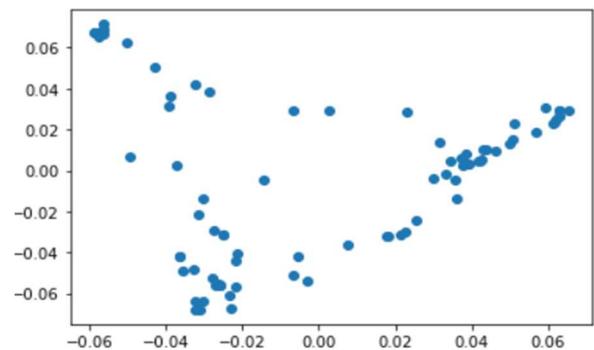


Fig. 5 Grouping of data attribute sets based on matrices that have similarities

C. Elbow Method

The selection of clusters in the K-Means algorithm is determined using the elbow method. This method calculates the maximum distance between the centroids using WCSS (within-cluster sum of squares).

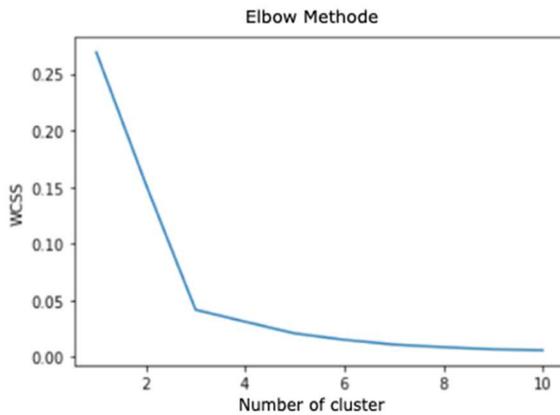


Fig. 6 Determination of the number of clusters with the elbow method

Figure 6 shows that the maximum number of clusters that can be employed is three. The number of clusters is calculated using the SSE (Sum Square Error) value or the inertia plot, which shows the SSE value on the Y axis and the number of clusters on the X axis.

D. K-Means Algorithm

The elbow method is used to determine the number of clusters in the K-Means Algorithm's clustering process. The K-Means algorithm seeks to reduce variation inside a cluster while increasing variation between clusters. It likewise seeks to locate groups in the data, with the variable k representing the number of groups (clusters). Figure 7 shows how the clustering is divided into categories based on the distance between the centroids and the data attributes.

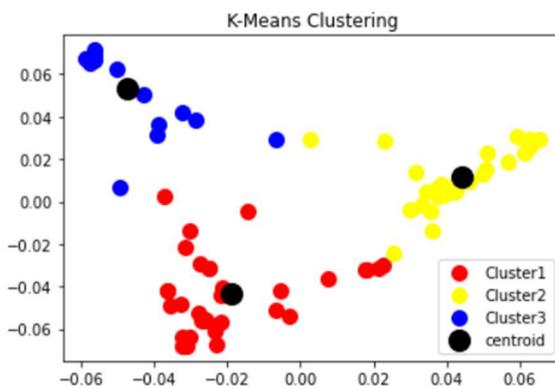


Fig. 7 Clustering process using K-Means algorithm.

ICFL participating students in this survey totaled 178 batches of 2017, 2018, and 2019. ICFL participating students in the Bachelor of Information Systems study program mostly took part in the type of Internship/Industrial Practice program, namely 42%, followed by independent studies and projects in villages, while the program with the lowest interest was ICFL Teaching in Schools. When viewed based on the program, most ICFL participating students in the Bachelor of Information Systems study program took part in the Certified

Internship program at 16.9%, followed by WRAP Internship and RISE 2021, while the program with the lowest interest was the KBMI which the Ministry of Education and Culture organized.

The clustering process using the K-Means algorithm, begins with determining the number of clusters through the elbow method. The K-Means algorithm aims to reduce variation inside a cluster while increasing variation between clusters. It also seeks to locate groups. The study's findings are based on the following steps of the Clustering process:

- Data Collection, which is carried out using the survey method;
- Data Selection is carried out using the required variables or attributes in research. Data is chosen by determining qualities from ICFL program participants' demographic data.
- Pre-processing/Cleaning: The 107 respondents' data is processed for the first time at this stage. Following the cleaning process, the number of data to be studied decreased from 107 to 83.
- Transformation: This stage generates a set of data ready for analysis.
- Data Clustering / Data Analysis.

E. Data Clustering Results

The K-Means algorithm is used to cluster the data, and the Python programming language is used to process it. Three clusters emerged as a result of the clustering, with the following criteria for each cluster:

TABLE I CLUSTER 1 CRITERIA	
GPA Range	2.29 to 3.97
Gender Proportion	Men (82%) Women (18%)
ICFL Program	ICFL - Rise (36%) ICFL - Internship (45%) ICFL - Independent Studies (12%) ICFL - IISMA (3%) ICFL - Teaching Campus (3%)
Admission path	Regular Academic Achievement (58%) Regular Phase Writing Examination (24%) Entrance Screening Examination (9%), General Admission (9%)
ICFL Category	Lectures (Student Exchange) (9%) Software Development (67%) Data Engineering (15%) Design and Analysis IS (3%) Teaching-Independent Campus (3%)
Year Semester	2020/2021 - Even semester (70%) 2021/2022 - Odd semester (30%)

TABLE II CLUSTER 2 CRITERIA	
GPA Range	2.71 to 3.89
Gender Proportion	Men (0%) Women (100%)
ICFL Program	ICFL - Rise (3%) ICFL - Internship (56%) ICFL - Independent Studies (28%) ICFL - IISMA (2%) ICFL - Teaching Campus (6%)
Admission path	Regular Academic Achievement (78%) Regular Phase Writing Examination (3%) Entrance Screening Examination (16%) General Admission (3%)
ICFL Category	Lectures (Student Exchange) (19%)

	Software Development (25%) Data Engineering (16%) Design and Analysis IS (34%) Teaching-Independent Campus (9%)
Year Semester	2020/2021 - Odd Semester (3%) 2020/2021 - Even Semester (9%) 2021/2022 - Odd Semester (88%)

TABLE III
CLUSTER 3 CRITERIA

GPA Range	2.57 to 3.85
Gender Proportion	Men (100%) Women (0%)
ICFL Program	ICFL - Rise (0%) ICFL - Internship (53%) ICFL - Independent Studies (41%) ICFL - IISMA (0%) ICFL - Teaching Campus (6%)
Admission path	Regular Academic Achievement (88%) Regular Phase Writing Examination (6%) Entrance Screening Examination (6%) General Admission (0%)
ICFL Category	Lectures (Student Exchange) (12%) Software Development (0%) Data Engineering (6%) Design and Analysis IS (65%) Teaching-Independent Campus (18%)
Year Semester	2020/2021 - Odd semester (9%) 2021/2022 - Odd semester (91%)

The resulting clustering shows that clusters 2 and 3 have a gender predominance in participating in the ICFL program. The cluster with the longest GPA range is Cluster 1, and the shortest is Cluster 2. Clusters 1 and 2 have 5 ICFL programs, and Cluster 3 has only three programs. Regular enrollment academic performance has the highest percentage of choice paths for the three clusters. The most visited ICFL category is Cluster 1, Software Development, with a share of 67%, and Clusters 2 and 3, Design and Analysis Information Systems. Semester 1 is dominated by 2020/2021; clusters 2 and 3, i.e., 2021/2022, are odd.

The existence of policies and research regarding ICFL in the lecture process at tertiary institutions affects the economic and social sectors. Among them is skill readiness in the world of work, both hard skills and soft skills. Based on a survey conducted through the SPADA DIKTI website regarding student opinion questions about the benefits of ICFL as a provision after graduation, it was found that both ICFL participants and non-participants agreed that the ICFL policy was very useful as a provision after graduation.

Based on the data, it was found that 63% of students generally stated that ICFL was very useful, while for ICFL participants there was an increase of 12%, namely 74% of students stated that ICFL was very useful. Thus, student participation in the ICFL program influences students' opinions about this program.

In addition, in this study, ICFL participants were also asked questions related to the skills they had acquired in the ICFL program. In the survey questions, students were able to choose several skills that were obtained according to the skill criteria at the Indonesian Assessment Center (ACI).

Based on these skills, an overall calculation was carried out on 133 ICFL participating students. The data obtained through the survey then identified which skills were dominantly obtained by ICFL participants, among the five

biggest, namely Creativity and Innovation, Conceptual Thinking, Adaptability, Strategic Management, and Execution Focused (Table IV).

TABLE IV
COMPETENCE OBTAINED FROM ICFL ACTIVITIES ACCORDING TO ICFL PARTICIPANTS

No	Competences	Sum of Students
1	Adaptability	103
2	Change Leadership	70
3	Conceptual Thinking	104
4	Creativity & Innovation	112
5	Customer Orientation	64
6	Entrepreneurship	51
7	Execution Focused	80
8	Fostering Teamwork	76
9	Impact & Influence	69
10	Nurturing People	40
11	Organizational Awareness	78
12	Strategic Management	81
13	Strategic Relationship	60
14	Concern For Order	55

The ICFL program provides benefits not only for students and the learning process in tertiary institutions but also positively impacts other sectors. The industrial sector, both state-owned and private companies, can interact directly with active students, who will be the output of higher education. Through this direct interaction, the industrial sector, among others, can prepare a work environment that is most suitable for the development of future generations. According to various studies, almost all industries have difficulty managing human resources across generations, including Baby Boomers and Generations X, Y, and especially Z and their successors. In addition, the industrial sector can also identify gaps between real needs in the world of work and students' educational curricula on campus. Industry can provide feedback to universities as part of the curriculum evaluation and improvement process. Of course, through this ICFL program, the industrial sector can also identify potential talents who, in the future, can become part of the company's progress.

For the government sector, the ICFL program can be synergized with programs from other ministries and agencies and other directorates general within the Ministry of Education and Culture. For example, the Merdeka Student Exchange and Teaching Campus programs allow students to get to know each other's culture and traditions in various archipelago regions. The same goes for the Young Warriors program, and students can contribute to social projects in various provinces. Apart from helping ministry programs, of course, the participation of students in the ICFL programs will strengthen their love for Indonesia.

IV. CONCLUSION

Based on the clustering result, it can be illustrated that gender dominance in participating in the ICFL program occurs in clusters 2 and 3, where in cluster 2 the proportion of female participants is 100% and in cluster 3 the proportion of male participants is 100%. The cluster with the longest GPA range is cluster 1, and the shortest is cluster 2. There are 5 ICFL programs in clusters 1 and 2, while in cluster 3 there are only 3 programs. The highest percentage of selection paths in the three clusters is JPA Regular. The most MBKM category

attended by students in cluster 1 is software development, with a proportion of 67%; clusters 2 and 3 are the design and analysis of information systems.

This study implements the K-Means algorithm for student demographic clustering based on the ICFL program. The data used focuses on 83 students and examines variables such as gender, GPA, university admission selection pathways, history of ICFL sub-program categories followed, and semester in which ICFL was taken. The resultant clusters can be categorized into three groupings based on the study conducted. The elbow method calculates the inertia, or the total value of the distance between each point and the nearest centroid, which is used to determine the number of clusters. The ICFL - Internship program is the most popular ICFL program among the three clusters, with over 40% of participants from each cluster participating. This demonstrates that the ICFL internship program is the most popular among students and is a valuable asset in starting a post-college career.

REFERENCES

- [1] Beritasatu.com, "MWA UI Holds Webinar Series on Economics to Health," 2021. <https://www.beritasatu.com/nasional/724771/mwa-ui-gelar-webinar-series-ekonomi-hingga-kesehatan>. (accessed Dec. 23, 2020).
- [2] Directorate General of Higher Education, *Independent Campus, Freedom to Learn Guide Book*. Jakarta: Directorate General of Higher Education, 2020.
- [3] S. Nurhayati and P. Puryati, "Assessment fairness in freedom of learning in open and distance education," *Research and Evaluation in Education*, vol. 8, no. 1, Jun. 2022, doi: 10.21831/reid.v8i1.45378.
- [4] Directorate General of Higher Education, "About Independent Campus Freedom to Learn," 2020. <https://kampusmerdeka.kemdikbud.go.id/web/about>. (Accessed Dec. 23, 2021).
- [5] A. Junaidi, *Panduan Penyusunan Kurikulum Pendidikan Tinggi Di Era Industri 4.0 Untuk Mendukung Merdeka Belajar-Kampus Merdeka*, IV. Jakarta: Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan, 2020.
- [6] K. Hartley and L. D. Bendixen, "Educational Research in the Internet Age: Examining the Role of Individual Characteristics," *Educational Researcher*, vol. 30, no. 9, pp. 22–26, 2001, doi: 10.3102/0013189X030009022.
- [7] D. R. Garrison, "Self-Directed Learning and Distance Education," in *Handbook of Distance Education*, M. G. Moore and W. G. Anderson, Eds. New Jersey: Lawrence Erlbaum Associates, Inc., 2003, pp. 161–168.
- [8] C. N. Gunawardena and M. S. McIsaac, "Distance Education," in *Handbook of Research on Educational Communications and Technology*, 2nd ed., D. Jonassen and M. Driscoll, Eds. New York: Lawrence Erlbaum Associates, Inc., 2004, pp. 355–395.
- [9] W. C. Diehl, "Charles A. Wedemeyer," in *Handbook of Distance Education*, Routledge, 2015. doi: 10.4324/9780203803738.ch3.
- [10] D. Ktoridou and E. Doukanari, "Task-based internships: Fostering ideal learning through focused experience," in *Proceedings of 2017 IEEE Global Engineering Education Conference, EDUCON 2017*, Jun. 2017, pp. 1717–1722. doi: 10.1109/EDUCON.2017.7943081.
- [11] P. C. Ooi, "Exploring ways to enhance students' internship experience," in *2016 IEEE 8th International Conference on Engineering Education (ICEED)*, 2016, pp. 107–108. doi: 10.1109/ICEED.2016.7856052.
- [12] S. Yamamoto, "The effect of an internship experience in Japan on Thai students' choice of employment," in *2018 5th International Conference on Business and Industrial Research (ICBIR)*, 2018, pp. 500–505. doi: 10.1109/ICBIR.2018.8391249.
- [13] G. Rulifson and A. Bielefeldt, "Influence of Internships on Engineering Students' Attitudes about Socially Responsible Engineering," in *2018 IEEE Frontiers in Education Conference (FIE)*, 2018, pp. 1–6. doi: 10.1109/FIE.2018.8658647.
- [14] F.-M. Sui, J.-C. Chang, and H.-C. Hsiao, "What Core Competence Can Students Learn from Off-Campus Internship?," in *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2020, pp. 640–644. doi: 10.1109/IEEM45057.2020.9309781.
- [15] M. Moltó-Aribau, A. T. Tarruella, and A. Vendrell-Vilanova, "Designing a Competency-Based Evaluation Model in External Academic Practices (Internship)," in *2018 IEEE Frontiers in Education Conference (FIE)*, 2018, pp. 1–4. doi: 10.1109/FIE.2018.8658675.
- [16] L. Andrade-Arenas and C. Sotomayor-Beltran, "On the Perspectives of Graduated Engineering Students on Three Dimensions of the Integrated Curriculum from a Peruvian University," in *2019 International Symposium on Engineering Accreditation and Education (ICACIT)*, 2019, pp. 1–4. doi: 10.1109/ICACIT46824.2019.9130268.
- [17] S. Chickerur, "Engineering Internships with Projects: A Workplace-Learning Model for Improving Student's Success and Performance," in *2015 8th International Conference on u- and e-Service, Science and Technology (UNESST)*, 2015, pp. 45–49. doi: 10.1109/UNESST.2015.21.
- [18] S. Andari, A. C. Setiawan, and A. Rifqi, "Student Exchange Program Of Merdeka Belajar-Kampus Merdeka (MBKM) In Covid-19 Pandemic," *Education and Learning Journal*, vol. 28, no. 1, pp. 30–37, 2021.
- [19] S. McKay, L. Lannegrand, J. Skues, and L. Wise, "Identity Development During Student Exchange: A Qualitative Study of Students' Perspectives on the Processes of Change," *Emerging Adulthood*, vol. 10, no. 2, pp. 420–433, 2022, doi: 10.1177/2167696820969464.
- [20] K. Govindasamy, A. V. Vidyapeetham, and V. Thambusamy, "Analysis of Student Academic Performance Using," *International Journal of Pure and Applied Mathematics*, vol. 119, no. November, pp. 309–323, 2018.
- [21] V. K. Anand, S. K. A. Rahiman, E. Ben George, and A. S. Huda, "Recursive clustering technique for students' performance evaluation in programming courses," in *2018 Majan International Conference (MIC)*, 2018, pp. 1–5. doi: 10.1109/MINTC.2018.8363153.
- [22] Z. Fan and S. Yan, "Clustering of College Students Based on Improved K-Means Algorithm," in *2016 International Computer Symposium (ICS)*, 2016, pp. 676–679. doi: 10.1109/ICS.2016.0139.
- [23] X. W. Su *et al.*, "A survey of attitudes to and demands on internship in graduates majoring in preventive medicine in 6 colleges and universities in China," *Zhonghua liu xing bing xue za zhi*, vol. 40, no. 7, p. 849–853, Jul. 2019, doi: 10.3760/cma.j.issn.0254-6450.2019.07.021.
- [24] I. Singh, A. S. Sabitha, and A. Bansal, "Student performance analysis using clustering algorithm," in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, 2016, pp. 294–299. doi: 10.1109/CONFLUENCE.2016.7508131.
- [25] D. Aggarwal and D. Sharma, "Application of Clustering for Student Result Analysis," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, no. 6C, 2019.
- [26] D. Kuswandi, E. Surahman, Z. Zufar At Thariq, and M. Muthmainnah, "K-Means Clustering of Student Perceptions on Project-Based Learning Model Application," in *2018 4th International Conference on Education and Technology (ICET)*, 2018, pp. 9–12. doi: 10.1109/ICEAT.2018.8693932.
- [27] I. Pasina, G. Bayram, W. Labib, A. Abdelhadi, and M. Nurunnabi, "Clustering students into groups according to their learning style," *MethodsX*, vol. 6, pp. 2189–2197, 2019, doi: <https://doi.org/10.1016/j.mex.2019.09.026>.
- [28] M. G. Vashisht, Bhawna, and A. Kalra, "Students Expectations using Opinion Mining Clustering," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9S, pp. 716–720, 2019.
- [29] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," *Knowl Eng Rev*, vol. 25, no. 2, pp. 137–166, 2010, doi: 10.1017/S0269888910000032.
- [30] P. Chapman *et al.*, *CRISP-DM 1.0: Step-by-step data mining guide*. 2000.