

Contributions of Data Science Educational Paradigm in a Disadvantages Area of Indonesia: a case study

Oviliani Yenty Yuliana^{1*}, Bernardo Nugroho Yahya², Remuz MB Kmurawak³

¹Sekolah Bisnis dan Manajemen, Universitas Kristen Petra

²Hankuk University of Foreign Studies

³Sistem Informasi Universitas Cenderawasih

Corresponding Author: Oviliani Yenty Yuliana oviliani@petra.ac.id

ARTICLE INFO

Keywords: Data Science, Data Analytics, Dashboard, Handling Missing Data, Machine Learning, Deep Learning

Received : 25, April

Revised : 27, Mei

Accepted: 29, June

©2023 Yuliana, Yahya, Kmurawak:

This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

The purpose of this community service was to disseminate knowledge on Data Science to lecturers and students enrolled in the Information Systems Program Study at Cenderawasih University. Through a webinar series, participants learned about the importance of data science and analytics. This initiative aims to integrate Data Science into the curriculum, which is currently in its final stages. The introduction of a Data Science course expects to enhance the competencies of human resources in Papua, supporting the developmental efforts of the Indonesian Government. Feedback from participants indicates their satisfaction with the speakers, fulfillment of expectations, and acquisition of new knowledge. Considering their input, the upcoming activity will organize as a workshop.

INTRODUCTION

In order to accelerate development progress, improve the quality of life, and narrow the development gap, among the 3T (Tertinggal, Terluar, dan Terdepan, it call Disadvantaged, Outermost, and Frontier) regions (Bunaiya, 2023; Setpres, 2022), the Government of Indonesia has introduced Presidential Regulation No. 131 of 2015. This regulation outlines the National Long-Term Development Plan for the period 2005-2025 (stipulated in Law No. 17 of 2007). The government actively promotes digital, green, and blue economic growth as part of these long-term development initiatives. This article focuses on the digital economy, where the Government of Indonesia implements regulations governing the implementation of electronic systems and transactions to increase global competitiveness. In addition, the government has established a legal framework for managing electronic transactions, payment activities, data protection, and information security in Indonesia.

Papua, one of the three main islands referred to as the 3T islands, has seen various initiatives by the Indonesian government to promote the growth of its digital economy. These measures include enhancing internet accessibility and digital infrastructure through the Palapa Ring program, which seeks to improve internet connectivity across all parts of Indonesia. The government also supports startups and technology by offering training, funding, and market access opportunities. In addition, incentives and simplified procedures for licensing and taxation facilitate digital businesses in Papua. Sufficient resources are essential to ensure the successful development of the digital economy in Papua.

The Indonesian Government has introduced commendable community service initiatives that contribute to advancing the digital economy. An excellent example is the provision of training and education to enhance the capabilities of human resources. By offering data science training, individuals in Papua can acquire improved skills and knowledge, effectively utilizing available technology and information to establish e-commerce ventures, enhance productivity, and make data-driven decisions. Consequently, this not only enhances welfare but also bolsters global competitiveness.

The right decision could be to offer Data Science training to students and lecturers at Cenderawasih University (UNCEN). The university is the oldest state university and one of the favorite campuses in Papua (Maulidina, 2021), as a place for community service. However, UNCEN must increase its exposure to technological advances and embrace digitalization. The Information Systems Study Program has obtained a B accreditation and is developing a curriculum with Data Science. As part of this effort, a lecturer from the study program invited lecturers from South Korea and Surabaya to share Data Science via Webinar. This community service's primary purpose is to provide students and lecturers with knowledge about Data Science. In addition, this service contributes to integrating Data Science into the course curriculum. Ultimately, these efforts indirectly assist the Government of Indonesia to prepare digital talent in Eastern Indonesia.

IMPLEMENTATION AND METHODS

Remuz MB Kmurawak, ST., MT. and Prof. Bernardo Nugroho Yahya, Ph.D. took the lead in introducing a program called TIK Talk, which stands for Information and Communication Technology (ICT) Talk. These activities aimed to offer valuable knowledge and perspectives to ICT stakeholders and students in Papua, focusing on the Information Systems program study within the Faculty of Mathematics and Natural Sciences at UNCEN. These activities comprised online sessions (webinar) featuring presentations and discussions conducted twice. Our intention with this community service initiative is to provide insights on implementing Data Science into the curriculum of the Information Systems program study at UNCEN, thus enhancing ICT proficiency.

The inaugural TIK Talk aimed to raise awareness regarding the significance of Data Science in various domains of life. This talk occurred on Monday, September 19, 2022, from 17:30 to 18:30 WIT. The distinguished speaker for this event was Prof. Bernardo Nugroho Yahya, Ph.D, as indicated in Figure 1. Notably, the first TIK Talk activity garnered considerable enthusiasm and witnessed the participation of 79 individuals, including both lecturers and students, as depicted in Figure 2.



Figure.1 The First TIK Talk Poster

The second TIK Talk webinar served as a follow-up to the previous session, focusing on the Data Science Series, explicitly delving into topics such as Data Analytics and Handling Missing Data. This talk occurred on Monday, October 17, 2022, from 17:30 to 18:30 WIT, as illustrated in Figure 3. Oviliani Yenty Yuliana, Ph.D., provided insights on Data Science, while Prof. Bernardo Nugroho Yahya, Ph.D., delivered the material on Handling Missing Data. Figure 4 features a photograph capturing the participants of this activity. Table 1 presents a summary of the details concerning the 48 participants. Most attendees from the Information Systems study program participated in the event, including lecturers, staff, students, and a Physics lecturer.

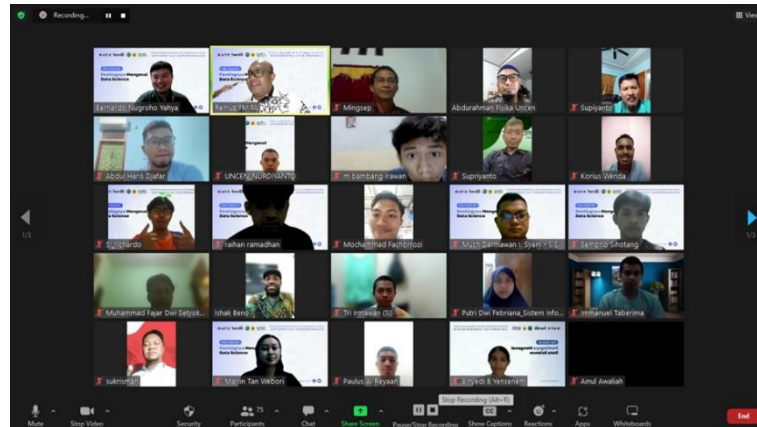


Figure.2 The First TIK Talk Participants



Figure.3 The Second TIK Talk Poster

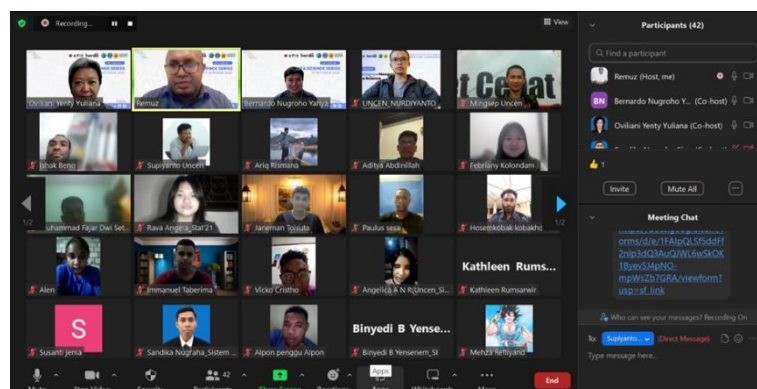


Figure.4 The Second TIK Talk Participants

Table.1 Details of the Second TIK Talk Participants

Department	Participant			Grand Total
	Lecturer	Staff	Student	
Information Systems	4	1	39	44
Mathematics	1		2	3
Physics	1			1
Grand Total	6	1	41	48

In the second TIK Talk webinar, participants completed a questionnaire that consisted of a five-point Likert Scale, i.e., 1: strongly disagree, 2: disagree, 3: neutral, 4: agree, and 5: strongly agree to gather their perceptions, insights, and feedback. The questionnaire included the following indicators: 1) the activities offered a fresh understanding of the subject matter, 2) the speakers demonstrated a strong command of the material presented, 3) the method employed in the activities was suitable for the topic, 4) The activities proved beneficial for the participants, and 5) the activities met the expectations of the participants. Furthermore, we evaluated the conducted community service activities.

RESULTS AND DISCUSSION

The first TIK Talk webinar discussed the importance of data science related to what data science and the importance of data science. Data science occupies a pivotal position at the intersection of various fields, including artificial intelligence (AI), machine learning (ML), deep learning (DL), big data, and data science itself. Data science extracts insights and knowledge from big data. It utilizes techniques and tools from AI to mimic human intelligence for making decisions and performing tasks that traditionally require human intervention. Machine learning creates algorithms that allow machines to learn from data and improve their performance over time without explicit programming. Deep learning leverages artificial neural networks to process and analyze complex patterns and structures in data. Big data plays an essential role in data science by providing vast data for analysis. In contrast, data science employs techniques to extract meaningful insights and knowledge from the data. In other words, data science is central for integrating AI, ML, DL, and big data to harness the power of data and drive innovation across various industries.

Furthermore, Data science is paramount in the 21st century because it can unlock valuable insights from the vast amounts of data generated daily. Using advanced analytics and machine learning techniques, data scientists can extract meaningful patterns, make informed predictions, and drive data-driven for decision-making across various industries. Data science is essential in solving complex problems, optimizing processes, and identifying opportunities for growth and innovation. In addition, the notion of data science being the “sexiest job” of the 21st century is still increasing demand for skilled data scientists with statistical expertise, programming skills, and domain knowledge. Lastly, the “data never sleeps” highlights the continuous generation and flow of data in our digital age, emphasizing the need for data scientists to harness the vast resource to gain valuable insights and stay forward in today’s data-driven world.

As Southeast Asia’s most prominent digital economy market, Indonesia faces numerous challenges requiring attention. In response to these opportunities and obstacles, the Indonesian government has established a strategic framework called the Indonesian National Strategy Framework for Artificial Intelligence (BPPT, 2020). The framework aims to guide by emphasizing the interconnections between four critical areas of focus: 1) Ethics and Policy, 2) Talent Development, 3) Infrastructure and Data, and 4) Industrial Research and Innovation. Additionally, this framework outlines specific priority areas, namely: 1) Health Services, 2) Bureaucratic Reform, 3) Education and Research, 4) Food Security, and 5) Mobility and Smart Cities. By utilizing artificial intelligence, the framework facilitates the integration of national initiatives from the four focus areas with the priority areas. Figure 5 illustrates Indonesia’s National Strategy Framework for Artificial Intelligence.

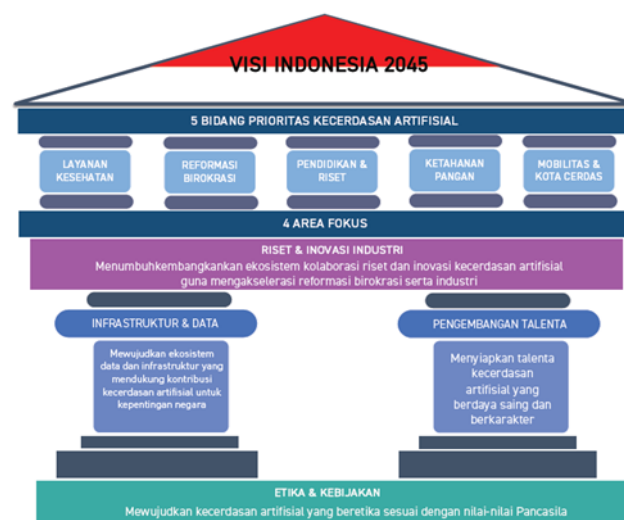


Figure.5 Indonesia’s National Artificial Intelligence Strategy Framework

The second TIK Talk webinar delivered the Missing Data topic. It covered two main issues, i.e., Understanding Missing Data which is a Challenge in AI, and Strategies for Managing Missing Data to address the significance of handling missing data. Analysts frequently encounter the challenge of addressing missing data arising from various factors, as shown in Figure 6. These factors encompass respondent omissions, refusals, or incomplete survey responses, as well as issues like sensor failures, intentional tool disabling, power outages, changes in data capture methods, lost internet connection, network failures, corrupt hard drives, and interrupted data transfers. Missing values fall into three categories: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR). MCAR denotes data missing independently of the dataset values, such as instrument faults or internet connection errors. MAR predicts missing data values based on other variables in the dataset, such as when data is missing for specific age groups or genders about connectivity in rural areas or income. MNAR represents the most intricate type of missing data, involving identification and handling challenges, such as missing data resulting from intentional actions of specific drivers or when the city is “Jayapura.”

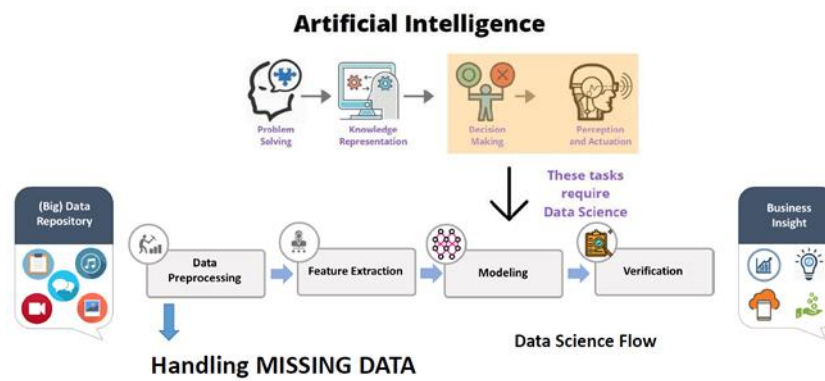


Figure.6 Data Science Flow

Handling missing data can be approached in various ways, as shown in Figure 7. One option is to employ deletions, which involve removing the tuples with missing values. It is typically when the class label is absent. However, this method may not be effective when the percentage of missing values per attribute significantly varies. Two common strategies for deletions are pairwise deletion, which only eliminates the missing values, and list-wise deletion, which deletes the entire rows containing missing values. Another deletion approach is column deletion to remove the columns with missing values. Secondly, fill in the missing values manually or statistically, particularly for non-time series data. A manual filing can be a laborious and impractical process. Statistical techniques offer more efficient alternatives. One approach is to fill missing values with a constant, such as labeling them as “unknown.” Alternatively, mean, median or mode values can be used to substitute missing values based on the statistical characteristics of the data. Lastly, employ advanced techniques to handle missing data. One such technique is the utilization of the kNN (k Nearest Neighbor) algorithm, which imputes missing values based on similar neighboring instances. Another method is Multiple Imputation by Chained Equations (MICE), which enables multivariate imputation by generating plausible values through an iterative process. Additionally, k-means clustering can help address missing data.

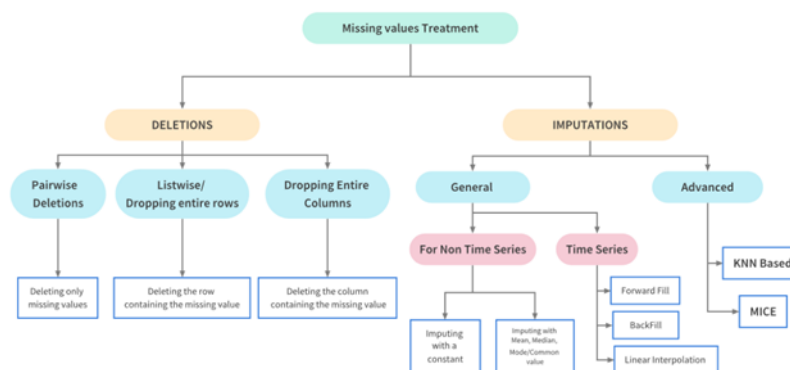


Figure.7 Missing Values Treatment

Another the second TIK Talk was Data Analytics Introduction, which discussed 1) motivating the need for digital talent as a data science to support the development of the Indonesian Government’s digital economy, 2) introducing three types of data/business analytics (descriptive, predictive, prescriptive)3) demonstrating visualization data analytics in the form of dashboards. In other words, business analytics and dashboards closely relate to analyzing and visualizing business data. Business analytics involves collecting, cleaning, and analyzing data to generate valuable insights. In contrast, dashboards are tools to present data visually that is easier to understand.

Sharda et al. (2019) and (Kunc & O’Brien, 2019) generally categorize Business Analytics into three main types of analytics such as Descriptive, Predictive, and Prescriptive, as illustrated in Figure 8. Descriptive Analytics, the first type, focuses on answering the questions “What happened?” and “What is happening?” It involves business reporting, dashboards, scorecards, and data warehousing. This kind of analysis helps identify well-defined business problems and opportunities. Predictive Analytics, the second type, aims to answer the questions “What will happen?” and “Why will it happen?” It enables techniques like data mining, text mining, web/media mining, and forecasting. Predictive analytics strives to generate accurate projections of future events and outcomes. The last type, Prescriptive Analytics, addresses the questions “What should I do?” and “Why should I do it?” It enables optimization, simulation, decision modeling, and expert systems. Prescriptive analytics aims to provide the most optimal business decisions and actions. In short, Sharda et al. (2019) divide Business Analytics into three types: descriptive analytics for understanding what happened and what is happening, predictive analytics for forecasting what will happen and why, and prescriptive analytics for determining what should be done and why.

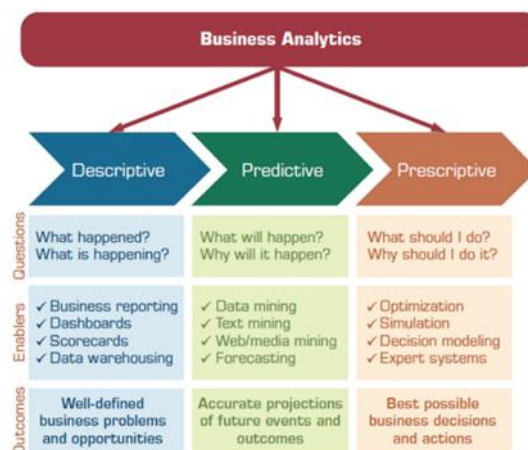


Figure.8 Business Analytics Framework
Source: (Sharda et al., 2019)

Business analytics, a specialized application of data science, focuses on leveraging analytical techniques and tools to gain insights into business operations, identify patterns and trends, and provide recommendations for enhanced decision-making (Lu et al., 2021). On the other hand, dashboards serve as visual interfaces that consolidate and present relevant information concisely and comprehensibly (Yalcin et al., 2022; Yap, 2020). For example, Figure 9 shows a Tokopedia sales dashboard. Dashboards integrate data science and business analytics outputs, offering real-time monitoring of key performance metrics, data visualizations, and crucial reports, enabling stakeholders to make informed decisions based on the synthesized information. The combination of data science, business analytics, and dashboards forms a cohesive framework that empowers organizations to unlock the value of data and gain a competitive edge in today's data-driven landscape.

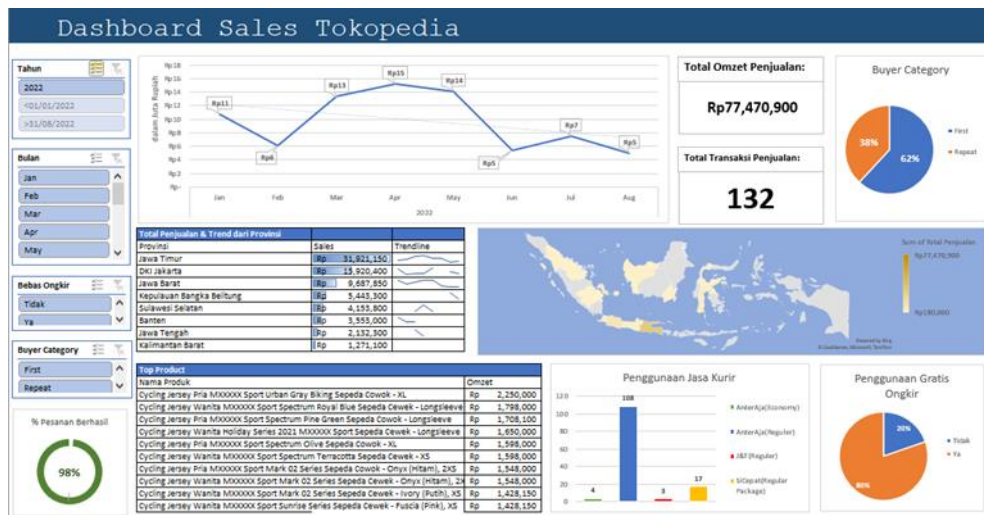


Figure 9. Dashboard Example

In the second ICT Talk activity, we distributed questionnaires to students. The results of the questionnaire are in Figure 10. In general, students (80%) strongly agreed, 18.05% agreed, and 1.95% were sufficient on all question indicators. Most students strongly agree (95.12%) and agree (4.88%) with the speakers' mastery of the material presented. They also strongly agree (78.05%) with the activity method providing a good topic and Activity to meet the participants' expectations. In addition, students also strongly agree (75%.61) and agree (24.39%) with activities that benefit the participants. Finally, students strongly agree (73.17%) and agree the highest of all indicators (26.83%) that the activity provides a new understanding of the topic.

Description	Strongly Disagree		Unagree		Neutral		Agree		Strongly Agree	
	Σ	%	Σ	%	Σ	%	Σ	%	Σ	%
Activities provide a new understanding of the topic	0	0	0	0	0	0	11	26.83	30	73.17
Speakers master the material presented	0	0	0	0	0	0	2	4.88	39	95.12
The activity method provides an adequate the topic	0	0	0	0	4	9.76	5	12.20	32	78.05
Activities are beneficial for the participants	0	0	0	0	0	0	10	24.39	31	75.61
Activities meet the expectations of the participants	0	0	0	0	0	0	9	21.95	32	78.05
Average							1.95		18.05	80.00

Figure.10 Student Satisfaction with the Second TIK Talk

$$class\ intervals = \frac{highest\ scale - lowest\ scale}{number\ of\ classes} \tag{1}$$

To assess student satisfaction with each assessment indicator, we compute the average perception of student satisfaction. Subsequently, we determine the class interval using the formula (1), which yields a value of 0.8, based on the calculation (5-1)/5. The classification scale is established as follows: ratings ≤ 1.8 indicate a state of being very dissatisfied, ≤ 2.6 represents dissatisfaction, 3.4 ≤ signifies a neutral response, ≤ 4.2 implies satisfaction, and ratings ≤ 5 indicate a state of being very satisfied. Figure 11 illustrates that the average student satisfaction perception ranges from 4.68 to 4.95 within the class interval above 4.2. In another word, all students expressed great satisfaction with the TIK Talk activities.

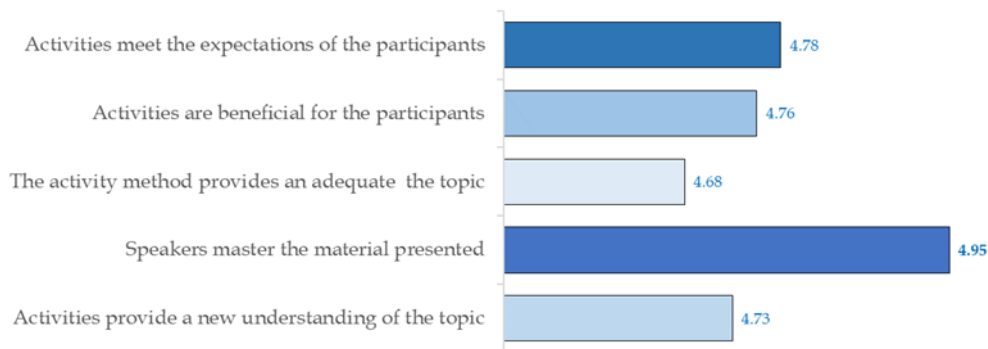


Figure.11 Summary of Student Satisfaction with the Second TIK Talk

Nevertheless, a neutral response was given by four students (9.76%), indicating that the activity-based approach offers a suitable subject. Furthermore, participants put forward several recommendations, including: 1) incorporating practical exercises and workshops to enhance participants' comprehension and 2) covering more current and pertinent topics like IoT, artificial intelligence, and computer vision.

CONCLUSIONS AND RECOMMENDATIONS

We conducted community service by organizing a series of TIK Talks sessions at UNCEN's Faculty of Mathematics and Natural Sciences, the Information Systems study program. These talks aim to emphasize the importance of data science, introduce data analytics, and provide guidance on handling missing data. The students enrolled in the program have expressed great satisfaction with these TIK Talk webinars, rating them highly between 4.68 and 4.95. As a result of the TIK Talk webinars held in 2022, the Information Systems study program is currently developing the final phase of its curriculum to incorporate the principles and practices of data science. Considering the feedback provided by the participants, the community service can expand its efforts by presenting the materials through workshops. This approach aims to enhance participants' understanding, enabling them to grasp contemporary and relevant topics, including IoT, artificial intelligence, and computer vision.

ACKNOWLEDGMENT

We are grateful to the Information Systems study program at UNCEN for allowing us to share data science and business analytics knowledge to enhance the curriculum. We also thank Kefas Alfando for developing a dashboard showcasing descriptive and exploratory analysis during the Introduction to Data Analytics webinar.

REFERENCES

- BPPT. (2020). *Strategi Nasional Kecerdasan Artifisial Indonesia 2020-2045*.
- Bunaiya. (2023). *President Jokowi is considered successful in building Indonesia evenly*.
<https://www.rri.co.id/index.php/nasional/130917/presiden-jokowi-dinilai-berhasil-bangun-indonesia-secara-merata>
- Kunc, M., & O'brien, F. A. (2019). The role of business analytics in supporting strategy processes: Opportunities and limitations. *Journal of the Operational Research Society*, 70(6), 974--985.
- Lu, J., Cairns, L., & Smith, L. (2021). Data science in the business environment: customer analytics case studies in SMEs. *Journal of Modelling in Management*, 16(2), 689--713.
- Maulidina, N. (2021). *The 5 Most Favorite Universities in Papua, Millennials Must Know*. <https://papua.inews.id/berita/universitas-di-papua>

- Setpres, B. (2022). *President Jokowi: Regional Expansion in Papua for Equitable Development*. <https://www.menpan.go.id/site/berita-terkini/dari-istana/presiden-jokowi-pemekaran-wilayah-di-papua-untuk-pemerataan-pembangunan>
- Sharda, R., Delen, D., Turban, E., & King, D. (2019). *Business Intelligence, Analytics, And Data Science: A Managerial Perspective* (4th ed.).
- Yalcin, A. S., Kilic, H. S., & Delen, D. (2022). The use of multi-criteria decision-making methods in business analytics: A comprehensive literature review. *Technological Forecasting and Social Change*, 174.
- Yap, A. Y. (2020). Creating Business Analytics Dashboard Designs Using Visualization Methodologies: Case Methods for Innovative Analytics Pedagogy. *Information Systems Education Journal*, 18(5), 25–33.