WORKSHOP JUNIOR DATA SCIENTIST

Mega Bagus Herlambang[1], Ni Made Sudri[2], Linda Theresia[3], Gadih Ranti[4], Yasmin Mauliddina[5]*, Silvia Merdikawati[6], Aditya Maulana[7], Dedes Bangkit Munawar[8], Nafia Rahmah[9]

[1,2,3,4,5,6,7,8,9)] Program Studi Teknik Industri, Institut Teknologi Indonesia
e-mail: yasmin.mauliddina@gmail.com

Abstract

The Junior Data Scientist Workshop, conducted on August 25, 2024, was designed to introduce participants to foundational concepts in Data Science and essential data analysis tools, with a focus on the Orange platform. Five students with a keen interest in data science attended the event. The workshop utilized a blend of theoretical instruction and practical application, guiding participants through key processes such as data cleaning, statistical analysis, and data visualization. Post-workshop evaluations indicate that 85% of participants demonstrated a marked improvement in their understanding of Data Science concepts and methods. Despite the participants' successful application of analysis techniques, several challenges were noted in comprehending complex statistical principles, suggesting a potential benefit from supplementary instructional sessions. This workshop has contributed significantly to advancing participants' data literacy and analytical competencies, equipping them with critical skills essential for data-driven decision-making.

Keywords: Data Science, Workshop, Data Analysis, Orange Platform, Improvement of Understanding, Analytical Skills.

INTRODUCTION

In today's rapidly evolving digital landscape, data-driven decision-making has become critical for organizations across diverse sectors, as they increasingly rely on data insights to shape strategic directions. Consequently, there is a rising demand for professionals equipped with the analytical skills to interpret and extract meaningful insights from complex datasets. The Junior Data Scientist Workshop is structured to provide participants with essential knowledge and hands-on skills in data science, preparing them to excel in this dynamic field.

This workshop targets aspiring data scientists who seek to enhance their data literacy and analytical acumen. Through a blend of theoretical knowledge and practical application, participants will explore key data analysis tools, with an emphasis on the Orange platform. Known for its user-friendly interface and robust capabilities in visual data analysis, Orange serves as an ideal tool for teaching data science fundamentals. The workshop will cover core components of the data analysis workflow, including the use of widgets for data manipulation and descriptive statistical techniques to examine dataset characteristics.

Participants will further develop their skills by engaging in regression and classification modeling exercises, which will demonstrate the application of analytical techniques in constructing predictive models. This approach aims to provide participants with a comprehensive understanding of the data analysis process and its practical applications in real-world settings. Ultimately, the workshop underscores the importance of data science competencies in supporting data-driven decision-making, equipping participants with the tools to meet the growing analytical demands of modern organizations.

METHOD

The Junior Data Scientist Workshop will adopt a structured and interactive approach to ensure participants gain a thorough and practical understanding of foundational data science concepts. The workshop methodology is organized into distinct, sequential phases designed to build a comprehensive learning experience.

Orientation and Introduction

The workshop will begin with an orientation session, introducing participants to fundamental data science concepts, including the strategic role of data in decision-making and its diverse applications across industries. This session will outline the workshop's objectives and expected outcomes, providing context and highlighting the value of data science skills in contemporary organizational settings.

Theoretical Foundation

Participants will engage in sessions covering theoretical foundations in data science, focused on key topics essential for analytical proficiency:

• **Data Collection**: This segment introduces techniques and tools for data acquisition from multiple sources, emphasizing distinctions between qualitative and quantitative data collection methods and their relevance in different analytical contexts.

• **Data Cleaning**: Participants will learn methods for data cleaning and preparation, emphasizing techniques to eliminate duplicates and irrelevant information. This process is critical to enhancing data quality, ensuring accurate and meaningful analysis Data Cleaning: Participants will learn methods for data cleaning and preparation, emphasizing techniques to eliminate duplicates and irrelevant information. This process is critical to enhancing data quality, ensuring accurate and meaningful analysis.

• **Data Analysis**: This section covers the principles of statistical analysis and introduces various data modeling techniques. Using the Orange platform's intuitive widget-based interface, participants will gain practical skills in processing datasets (such as CSV files) and applying classification models to extract insights from data.

• **Data Visualization**: The final theoretical component will focus on the interpretation and presentation of analysis results. Participants will learn to use data visualization tools to transform analytical findings into insights that can support data-driven decision-making.

Direct Practice

Following the theoretical sessions, participants will engage in applied exercises to reinforce their understanding of data analysis principles through hands-on practice with a provided dataset. This practical segment includes:

• **Utilization of Data Analysis Tools:** Participants will work with Orange, a user-friendly data analysis and visualization software, to gain experience in navigating analytical tools and software environments commonly used in data science.

• **Application of Data Cleaning and Analysis Techniques:** Participants will apply data preparation techniques to enhance data quality, followed by statistical and analytical methods to extract insights, enabling them to perform end-to-end data processing.

• **Development of Data Visualization:** To effectively communicate their findings, participants will construct visual representations of the analysis results, emphasizing the importance of data visualization in interpreting and presenting data-driven insights.

This hands-on component aims to deepen participants' technical skills, bridging theoretical knowledge with practical application to prepare them for real-world data analysis scenarios.

Mentoring and Discussion

During the practical sessions, instructors and mentors will actively guide participants, offering individualized support to enhance their learning experience. Interactive discussions will be conducted to address any challenges encountered, allowing participants to collaboratively explore solutions and deepen their understanding. This approach not only fosters peer learning but also reinforces problem-solving skills, making the practical application of concepts more effective and meaningful.

Fig. 1. Workshop in Progress

Fig. 2. Presentation by Instructor

Fig. 3. Workshop Concluded

Evaluation and Conclusion

The workshop will conclude with an evaluation session, enabling participants to reflect on and provide feedback regarding their learning outcomes. This feedback will be systematically gathered and analyzed to inform enhancements for future workshops, ensuring continuous improvement in workshop quality and relevance. Additionally, participants will receive certificates as formal recognition of their engagement and successful completion of the workshop, marking their achievement in foundational data science skills.

RESULT AND DISCUSSION

The results and discussion section presents the findings from the workshop and provides an in-depth analysis of participants' learning outcomes. This section aims to examine the effectiveness of the workshop in enhancing participants' understanding of data science concepts, as well as to identify challenges and areas for improvement. Through a detailed exploration of the feedback and performance data, we aim to draw insights that will inform future iterations of the workshop and contribute to the advancement of data science education.

Improvement in Understanding Data Science Concepts

Before the workshop, participants demonstrated a limited understanding of Data Science. However, following the introductory session and hands-on practice, 85% of participants reported a significant improvement in their comprehension of fundamental Data Science concepts, including key definitions, objectives, and industry applications. This outcome highlights the effectiveness of a practice-based learning approach in deepening participants' understanding and reinforces the value of experiential learning in the domain of Data Science.

Table 1 List of Participants

Name	Institution	Pre-test Score	Stage-3 results
Ilham Ikram Ramdhan	SMK Taruna baja	60/100	90 / 100
Muhamad Sandi Alfaizi	SMA Alfityan School Tangerang	90 / 100	90 / 100
Khiar Zaki Maulana	SMA Riyadhussholihiin Pandeglang	90 / 100	90 / 100
Ahmad Abdul Dzaki	SMKN 7 Kabupaten Tangerang	40 / 100	90 / 100
Muhammad Alfian	SMKN 7 Kabupaten Tangerang	40 / 100	100 / 100

Application of the Orange Data Analysis Tool

During the workshop, participants were trained in the use of the Orange platform for data analysis, with a focus on key features such as data cleaning, visualization, and the application of classification algorithms. Following the practical session, participants expressed a high level of comfort in using the platform. This suggests that the intuitive, visual interface of Orange plays a crucial role in facilitating the understanding of data analysis processes, particularly for beginners, by reducing the need for advanced programming knowledge. This ease of use supports the platform's effectiveness in bridging the gap between theoretical learning and practical application in data science.

Data Analysis and Model Application

Participants conducted data analysis using the provided dataset, with all successfully identifying patterns and trends, as well as applying classification models to predict specific outcomes. The analysis revealed that the utilization of certain features within the application significantly contributed to enhancing user satisfaction, which became the focal point of the discussion. These findings underscore the practical application of data analysis techniques in deriving actionable insights and demonstrate the effectiveness of the participants' analytical skills in addressing real-world problems.

Challenge and Alternatives

While many participants demonstrated notable progress, several challenges were identified. Some participants experienced difficulty grasping the statistical concepts underlying data analysis. To address this, it is recommended that future workshops include an additional session focused on the fundamentals of statistics and data analysis. Furthermore, participants suggested that more time be allocated for hands-on practice, allowing them to explore the tools in greater depth and reinforce their practical skills. These adjustments could enhance the overall learning experience and better support participants in mastering key concepts and techniques

CONCLUSION

The Junior Data Scientist workshop, held on August 25, 2024, effectively enhanced participants' understanding of foundational Data Science concepts and the application of data analysis tools, particularly the Orange platform. Through a structured and interactive approach, 85% of participants reported an improvement in their knowledge following the introductory session and hands-on exercises.

The workshop also highlighted the efficacy of intuitive data analysis tools, such as Orange, which enable participants to perform data analysis without requiring extensive programming expertise. While participants demonstrated proficiency in applying analysis and classification techniques, some challenges emerged regarding the understanding of the statistical concepts underlying data analysis. In response, it is recommended that future workshops incorporate additional sessions focusing on the

fundamentals of statistics, along with extended hands-on practice to deepen participants' practical skills.

Overall, the workshop contributed positively to advancing participants' data literacy and analytical capabilities, essential competencies in the context of data-driven decision-making. These findings underscore the importance of integrating both theoretical instruction and practical application in data science education.

REFERENCES

Biecek, P., & Burzykowski, T. (2019). Explanatory model analysis: A new approach to data science. Springer.

Camacho, J., Pérez-Villegas, A., & Ferrer, A. (2018). Machine learning for quality monitoring: Challenges and opportunities in the Industry 4.0 era. IFAC-PapersOnLine, 51(11), 641–646.

Gantz, J. F., & Reinsel, D. (2011). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. IDC iView.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Kelleher, J. D., & Tierney, B. (2018). Data science: A practical introduction to data science. The MIT Press.

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18–23.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Journal of Big Data, 1(1), 1–9.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. Journal of Business Logistics, 34(2), 77–84.

Witten, I. H., Frank, E., & Hall, M. A. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.

hong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of Industry 4.0: A review. Engineering, 3(5), 616–630.