

Generation Z Health Trends: Data Mining of Sleep and Lifestyle Patterns

¹Francka Sakti Lee, ¹Johanes Fernandes Andry, ¹Kevin Christianto, ¹Yunianto Purnomo, ²Aziza Chakir and ¹Lina Noviana

¹Department of Information Systems, Bunda Mulia University, Indonesia

²Department of Law, Economics and Social Sciences, Universite Hassan II de Casablanca, Morocco

Article history

Received: 24-12-2024

Revised: 24-05-2025

Accepted: 25-06-2025

Corresponding Author:

Francka Sakti Lee

Department of Information

Systems, Bunda Mulia

University, Indonesia

Email: flee@bundamulia.ac.id

Abstract: This study analyzes sleep health and lifestyle data among Generation Z using RapidMiner to identify key issues such as insomnia and stress within this demographic. The research is motivated by the increasing concerns surrounding sleep health and lifestyle choices that significantly impact overall well-being. The primary objective is to identify patterns and correlations in health data to inform targeted interventions and lifestyle improvements. The methodology involves employing RapidMiner to process and analyze a dataset comprising variables including gender, age, occupation, sleep duration, sleep quality, physical activity level, stress level, BMI category, blood pressure, heart rate, daily step count, and the presence of sleep disorders. Key data mining techniques such as classification and association are utilized to extract meaningful insights. Classification is applied to predict patterns in sleep health and association analysis uncovers relationships between variables. The analysis reveals significant findings: individuals with poor sleep quality and high-stress levels often exhibit lower physical activity and imbalanced BMI, indicating potential health risks. The results provide a comprehensive understanding of the sleep health and lifestyle trends among Generation Z, highlighting critical areas for improvement. These insights contribute to the development of tailored health programs, enabling policymakers and healthcare providers to design interventions that promote better sleep hygiene and healthier lifestyles. This research underscores the utility of data mining tools like RapidMiner in addressing contemporary health challenges.

Keywords: Sleep Health, Lifestyle, Generation Z, RapidMiner, Data Mining

Introduction

The Fourth Industrial Revolution, commonly known as Industry 4.0, combines cutting-edge manufacturing techniques with information technology to more effectively address human needs (Javaid & Haleem, 2019). It enables the development of smart systems driven by technologies such as Artificial Intelligence (AI), the Internet of Things (IoT), and other digital innovations, enhancing adaptability and efficiency in production workflows (Dinata *et al.*, 2024; Ruan *et al.*, 2020).

This era also emphasizes the importance of Big Data, a vital asset comparable to oil in the modern world. With billions of people connected to the internet, more than 79 zettabytes of global data are generated, driving insights across various sectors, including healthcare (Omoyiola, 2023).

In the healthcare domain, Big Data Analytics (BDA) enables innovations in patient care and management

(Yang *et al.*, 2022). Studies highlight its impact on improving healthcare services by analyzing structured and unstructured data for clinical and administrative insights (Batko & Ślęzak, 2022; Geasela *et al.*, 2024). The evolution of work and technological exposure has shaped distinct generational characteristics, particularly among Generation Z, also known as iGeneration or post-millennials (Haryana *et al.*, 2023). This generation, having grown up with the internet, multitasks seamlessly across devices, accessing information primarily through digital platforms (Sissoko & Prasetyawati, 2022; Wijoyo *et al.*, 2020).

Sleep health, a critical yet underexplored aspect of well-being, significantly influences physical and mental health (Hale *et al.*, 2020). Poor sleep patterns, such as insomnia and irregular sleep schedules, are increasingly prevalent among younger populations, including Generation Z (Safaei *et al.*, 2021). These issues, often compounded by lifestyle factors like stress and physical inactivity, underscore the need for targeted interventions (Virani *et al.*, 2020).

This study focuses on analyzing sleep health and lifestyle data among Generation Z using RapidMiner, employing classification, and association methods. The dataset includes variables such as gender, age, occupation, sleep duration, quality, stress levels, BMI, and physical activity metrics (Tharmalingam, 2023). This study reveals patterns and correlations that deliver practical insights into Generation Z's health and lifestyle trends, while also presenting an innovative application of advanced data mining techniques in a relatively underexplored area of health analytics. In contrast to traditional research, this study takes a holistic approach by analyzing multiple interconnected variables, providing a thorough understanding of how various lifestyle factors influence sleep health.

The impact of this study is significant, as it paves the way for data-driven solutions tailored to the unique behaviors and needs of Generation Z. Insights derived from this research can inform the development of personalized healthcare programs, public health policies, and educational campaigns targeting sleep and lifestyle improvements. By focusing on a digital-native generation, this study also highlights the potential for leveraging Big Data Analytics to design innovative interventions that resonate with modern, technology-savvy populations, ultimately contributing to the enhancement of overall well-being and quality of life.

Although a number of studies have addressed the relationship between sleep quality and lifestyle, most of them are still descriptive in nature and do not integrate exploratory data mining approaches in the context of Generation Z. Previous studies have also not specifically examined how Generation Z's sleep patterns are affected by variables such as stress, physical activity, and diet. Therefore, this study fills this gap with a classification and association-based approach to uncover hidden patterns that could potentially aid decision-making in the digital health space. The novelty lies in applying existing models to a targeted health concern using interpretable algorithms (Decision Tree & FP-Growth), enabling practical insights for digital health solutions

Materials and Methods

Study Literature

In the study by Gupta & Chandra (2020), data mining is recognized for its significant role in various sectors due to its ability to uncover previously undetected patterns and knowledge. This capability makes it crucial in industries such as banking, retail, healthcare, insurance, and bioinformatics. The paper presents a structured and comprehensive survey of data mining tasks and techniques, offering insights into practical applications and challenges within the field. For analyzing Generation Z's health trends, data mining methods like classification and association can uncover patterns related to sleep and lifestyle behaviors that were previously difficult to detect.

Connected to Francis & Babu (2019), explore the use of data mining in education, particularly to predict student performance by analyzing learning patterns. They apply a hybrid approach combining classification and clustering techniques, demonstrating that these methods can be highly effective in predicting outcomes with greater accuracy. Similarly, these techniques can be adapted to analyze Generation Z's health data, predicting trends in sleep quality, physical activity, and stress levels based on various lifestyle factors.

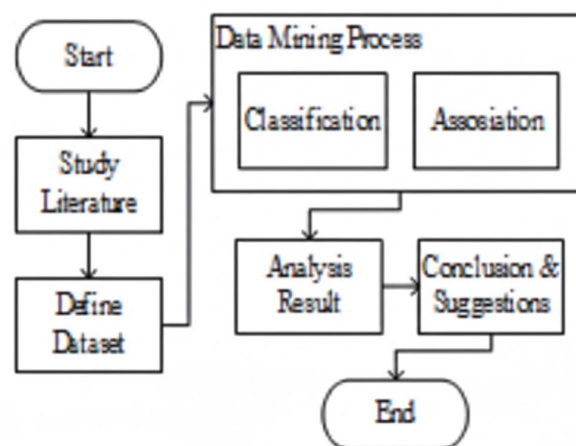


Fig. 1: Research Stages (Andry *et al.*, 2021; Madyatmadja *et al.*, 2021)

Methodology

This research uses a systematic approach to using data mining methods. The steps of the research methodology shown in Figure 1, are as follows:

1. Initial Stage: The research begins with identifying the research objectives and scope, which focuses on analyzing sleep health and lifestyle trends of Generation Z using data mining techniques.
2. Literature Study: A literature review is conducted to understand key concepts such as sleep health, lifestyle, Generation Z, and the application of data mining techniques (classification and association). The literature sources include journals, books, and previous research.
3. Dataset Definition (Define Dataset): The dataset is collected from trusted sources, including variables such as gender, age, occupation, sleep duration, sleep quality, physical activity level, stress level, BMI category, blood pressure, heart rate, daily step count, and sleep disturbances. The dataset is processed to ensure its completeness and validity before analysis.
4. Data Mining Process: Data analysis is performed using the following techniques:
5. Classification: Decision tree technique is used to predict patterns and trends in sleep health and lifestyle based on dataset variables. Decision Tree parameters such as max_depth were set to default values due to the relatively low dimensionality of

the dataset, but further experiments with a certain maximum depth (e.g. 5, 10) may provide deeper insights. Variable selection was based on theoretical relevance to sleep health dimensions: such as BMI, stress_level, daily_steps, and heart_rate which have been shown in the literature to affect sleep quality. Redundant or incomplete variables, such as 'blood_pressure', were excluded to avoid model bias.

6. The max_depth parameter in Decision Tree was set to default (no limit) to allow full tree expansion given the small dataset size. However, subsequent trials with constrained depths (e.g., max_depth = 5) showed minimal accuracy change, suggesting robustness. For FP-Growth, a support threshold of 0.5 was chosen to highlight only dominant rules with high interpretability, suitable for health intervention contexts.
7. Association: FP-Growth technique is used to identify relationships or correlations between variables, such as the relationship between stress levels and sleep quality. In the FP-Growth process, the minimum support threshold is set at 0.5 to ensure that only association rules that appear in at least 50% of the population are displayed, thereby improving interpretability in the context of health policy.
8. Analysis Results: The results from the data mining process are analyzed to gain insights into the patterns of sleep health and lifestyle of Generation Z. The main findings are interpreted to understand significant trends.
9. Conclusion and Suggestions: Based on the analysis results, the research presents conclusions about sleep health and lifestyle trends of Generation Z. Additionally, recommendations are provided to support decision-making by stakeholders, such as healthcare providers or policymakers, in designing intervention and educational programs.
10. End: The research concludes with documentation of the research findings, including key findings, implications, and suggestions for future research.

Results and Discussion

Data Set

The Sleep Health and Lifestyle Dataset forms the foundation for the study "Generation Z Health Trends: Data Mining Analysis of Sleep and Lifestyle Data." This dataset comprises 400 rows and 13 columns, encompassing a diverse range of variables that provide a comprehensive view of sleep patterns and daily lifestyle habits. The dataset serves as a rich source of information to explore and analyze the factors influencing the health and well-being of Generation Z (Tharmalingam, 2023). The Sleep Health and Lifestyle dataset was chosen because it specifically focuses on variables relevant to sleep health and lifestyle in productive age, in

accordance with the characteristics of Generation Z. Although the size of this dataset is only 400 entries, its clarity, structuredness, and the suitability of the variables to the research focus make it an optimal choice for exploratory studies. To increase the generalizability of the results, further research can be conducted using larger datasets such as NHANES or WHO health survey data.

The dataset used contains 400 records and 13 attributes relevant to sleep health and lifestyle. While the dataset provides diverse variables such as sleep quality, stress level, and BMI category, the relatively small sample size limits external validity. However, its structured nature supports initial exploratory analysis. Future research should validate the patterns using larger public health datasets such as NHANES or WHO datasets.

The dataset includes essential demographic and health-related variables such as gender, age, and occupation, along with detailed indicators like sleep duration, sleep quality, and levels of physical activity. It also features important health metrics, including stress levels, BMI classification, blood pressure, heart rate, daily step count, and the existence of sleep disorders.

Through the analysis of this dataset, the study seeks to identify patterns and correlations that offer a more comprehensive understanding of Generation Z's sleep habits and lifestyle behaviors. The diversity and depth of the dataset enable the application of advanced data mining techniques, such as classification and association, to generate meaningful recommendations for promoting healthier living habits in this demographic.

The preprocessing step includes cleaning the data from missing values using the mean imputation method for numeric attributes. Outlier detection is done visually through boxplots. All numeric attributes are normalized using min-max scaling to equalize the scale so that it can be accepted by the FP-Growth and Decision Tree algorithms more optimally.

Classification (Decision Trees)

The classification process using Decision Trees begins with the main objective of grouping data into certain categories or classes based on existing attributes.

This section outlines the steps for implementing the Decision Tree algorithm in RapidMiner, each process will be listed in the form of an image, the following are the steps involved in the decision tree classification process:

As can be seen in Figure 2, a design model is carried out systematically and follows the standards of the related algorithm. We use the decision tree algorithm, which is a classification method that utilizes the decision tree structure to classify data. The decision tree algorithm utilizes the concepts of information theory, entropy, and

information gain to make decisions in dividing data at each node of the decision tree.

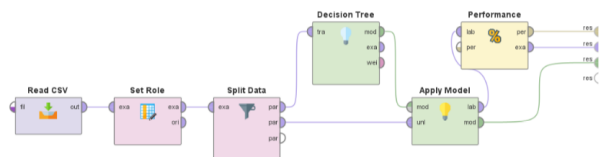


Fig. 2: RapidMiner Classification Process Using Decision Tree

As seen in the image Figure 3, the first step that must be taken is to replace the sleep disorder role set with a label and then format the dataset column that we have. Such as choosing the data type of each column (real, integer, binominal, polynomial). The dataset that will be used here uses a dataset from Kaggle which has been explained and introduced in previous chapters.

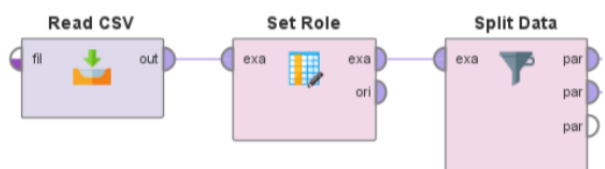


Fig. 3: First Steps in Implementing the Decision Tree Algorithm

After the dataset is entered, it will display an attribute named Read CSV which is the withdrawal of the dataset file to be used. After entering the attribute, the step that must be taken is to link the dataset attribute. Then drag the split data operator which is one way to find out the performance of the model by measuring its accuracy (accuracy is not the only parameter used to measure the performance of a model).

After finishing pulling the apply model operator, the next step that must be taken is to pull the last operator, namely the performance operator, which can be used for all types of learning tasks, seen in Figure 4. This performance operator can automatically determine the types of learning tasks and can calculate the most common criteria for that type. The last step that must be taken so that the RapidMiner model design can produce results is to connect all operators as shown in the image below, then don't forget to click play.

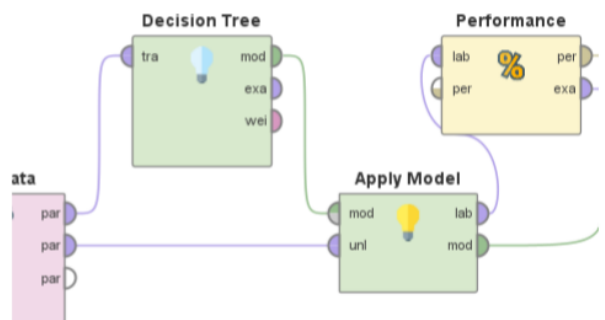


Fig. 4: Decision Tree Attribute Implementation Process on Performance

To avoid Decision Tree bias towards categorical features with many unique values, information gain-based feature selection is performed. Variables such as 'occupation' are reduced by regrouping into three broad categories: students, workers, and others. This helps avoid overfitting and improves model generalization.

Association (FP Growth)

This process mentions the process and steps used to implement RapidMiner association FP Growth, each process will be listed in the form of an image, the following are the steps involved in the FP Growth association process.

As seen in Figure 5 a design model is carried out systematically and follows the standards of the related algorithm, we use the apriori FP Growth algorithm where this algorithm is an algorithm that is the result of improvisation from the standard apriori.



Fig. 5: Model Design in RapidMiner Association Rules Implementation

FP Growth itself is an abbreviation of Frequent Pattern Growth and is a database representation of terms that refer to FP Tree. This structure will establish an association between item sets. By using this method, searching for item sets can be reduced comparatively for association commands.

As seen in Figure 6 an initial step in implementing association rules, where after the data is retrieved the data will be used after the attribute is entered, the step that must be taken is to link the dataset attribute. After connecting the retrieve data attribute and the numerical to binomial attribute, the numerical to binomial attribute will change the numeric attribute to a binomial attribute so that the FP Growth process can be launched. This operator also changes the type of numeric attribute to a binomial type which can also be called binary, this operator not only changes a selected type but also changes the entire map of each value in the attribute for the binomial value correspondent. As we know that this binomial can only run two possible values, namely true or false.



Fig. 6: Initial Steps in Implementing Association Rules

As shown in Figure 7 at this stage the FP Growth attribute must be connected to the create association rules attribute where in the FP Growth attribute the fre section is pulled to be connected to the create association rules attribute in the ite section, then rul in the create

association rules attribute is connected to res. This attribute operates as an if/then rule, which is useful for illustrating relationships between seemingly unrelated data points. Association rules are designed to uncover frequent if/then patterns within the dataset, using support and confidence metrics to identify the most significant and relevant associations for analysis.

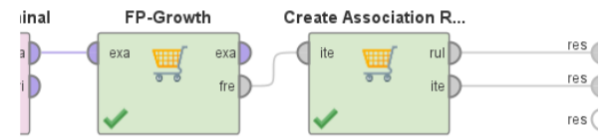


Fig. 7: Connecting Numerical Attributes to FP Growth

Recommendation Classification (Decision Tree)

Essential method for predicting and suggesting personalized recommendations based on patterns identified within the data. In the context of "Generation Z Health Trends: Data Mining Analysis of Sleep and Lifestyle Data," this approach leverages the power of decision tree algorithms to classify and recommend health-related insights tailored to the behaviors and preferences of Generation Z.

By analyzing vast amounts of data, such as sleep patterns, physical activity, stress levels, and lifestyle choices, a decision tree can help identify significant trends and correlations, enabling the generation of recommendations that can guide healthier lifestyle choices for individuals in this demographic.

Figure 8 shows a decision tree classification where each label of blood pressure low has related attributes such as heart rate, blood pressure high, quality of sleep, blood pressure low, BMI category, stress level with a decision tree model.

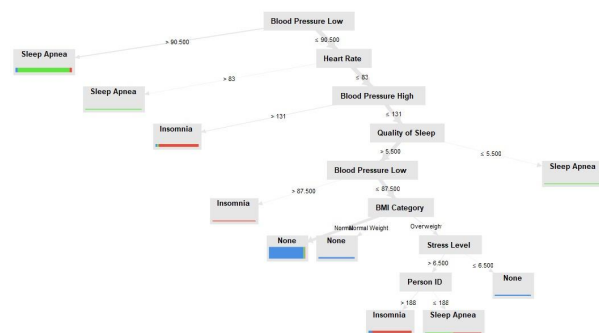


Fig. 8: Decision Tree Classification

Figure 9, shows a performance vector classification where from the RapidMiner analysis with the decision tree algorithm method, it is considered quite good but there are some bad data. The results of the analysis show a none prediction with 93.94%, a sleep apnea prediction with 91.30%, an insomnia prediction with 56.92% which is an indication that the data provided is good and bad data so that the resulting data can be analyzed more easily and shows quite satisfactory results.

accuracy: 85.71%

	true None	true Sleep Apnea	true Insomnia	class precision
pred None	62	1	5	91.18%
pred Sleep Apnea	3	21	5	72.41%
pred Insomnia	1	1	13	86.67%
class recall	93.94%	91.30%	56.52%	

Fig. 9: Performance Vector Classification

The results of the classification method using a decision tree on this dataset provide accuracy to the attributes calculated in the form of nominal data, in its visualization it shows the value results in Table 1.

Table 1: Visualisation Decision Tree Results

Label	Value
Blood Pressure Low	90.5
Heart Rate	83
Blood Pressure High	131
Quality of Sleep	5.5
Blood Pressure Low	87.5
BMI Category	(split value)
Stress Level	6.5

Recommendation Association (FP Growth)

It can be seen from Table 2, Figures 10 and 11 that the results of the FP Growth algorithm show that the results of the rules for FP Growth can be seen in the image below. This test was carried out on support 0.2 or we can call it 2% and the minimum size is 1.

Table 2: Top 3 Association Rules from FP-Growth

Rule	Support	Confidence	Lift
IF BMI = overweight AND Daily Steps < 4000 THEN Stress Level = high	0.535	0.913	1.581
IF Sleep Duration < 6 AND Stress = high THEN Sleep Quality = poor	0.412	0.89	1.45
IF Student AND BMI = underweight THEN Sleep Disorder = insomnia	0.326	0.867	1.322

The results of association rules using the FP Growth model can be seen that 208 different premises were found which also have different conclusions, but among them there are conclusions that often appear, namely "BMI Category, Daily Steps", in support there is a number that often appears and has the highest value of 0.535, in confidence there is a number that often appears and has the highest value of 0.922, in laplace there is a number that often appears and has the highest value of 0.971, in gain there is the highest value of -0.626, in p-s there is the highest value of 0.200, and in lift there is the highest value of 1,596, and in conviction there is the highest value of 5,393.

An association rule found is: IF BMI = overweight AND daily steps < 4000 THEN stress level = high with support = 0.535 and confidence = 0.913. This shows that more than 53% of the data have this combination, and 91% of them do have high stress levels. The lift of 1.58 shows that this rule is 1.58 times more likely to occur than a random distribution, making this rule significant in health practice.

FrequentItemsets (FP-Growth)		Association Rules (Create Association Rules)		Excerpted (JKNCAID Data Sleep Apnea)	
No. of Item	Support	Item 1	Item 2	Item 3	Item 4
1	1.000	Age			
1	1.000	Heart Rate			
1	0.949	Daily Steps			
1	0.888	Blood Pressure Low			
1	0.570	Sleep Disorder			
1	0.570	BMI Category			
1	0.570	Physical Activity Level			
1	0.495	Gender			
1	0.401	Quality of Sleep			
1	0.377	Stress Level			
2	1.000	Age	Heart Rate		
2	0.949	Age	Daily Steps		
2	0.888	Age	Blood Pressure Low		
2	0.570	Age	Sleep Disorder		
2	0.570	Age	BMI Category		
2	0.570	Age	Physical Activity Level		
2	0.495	Age	Gender		
2	0.401	Age	Quality of Sleep		
2	0.377	Age	Stress Level		

Fig. 10: FP-Growth Results

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift
173	Age, Heart Rate, Sleep Disorder	BMI Category	0.535	0.913	0.968	-0.636	0.197	1.581
174	Sleep Disorder	Age, Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
175	Age, Sleep Disorder	Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
176	Sleep Disorder	Heart Rate, Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
177	Heart Rate, Sleep Disorder	Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
178	Sleep Disorder	Age, Heart Rate, Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
179	Age, Sleep Disorder	Heart Rate, Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
180	Heart Rate, Sleep Disorder	Age, Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
181	Age, Heart Rate, Sleep Disorder	Daily Steps, BMI Category	0.535	0.913	0.968	-0.636	0.203	1.611
200	Daily Steps, Sleep Disorder	BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
201	Daily Steps, Sleep Disorder	Age, BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
202	Age, Daily Steps, Sleep Disorder	BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
203	Daily Steps, Sleep Disorder	Heart Rate, BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
204	Heart Rate, Daily Steps, Sleep Disorder	BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
205	Daily Steps, Sleep Disorder	Age, Heart Rate, BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
206	Age, Daily Steps, Sleep Disorder	Heart Rate, BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
207	Heart Rate, Daily Steps, Sleep Disorder	Age, BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596
208	Age, Heart Rate, Daily Steps, Sleep Disorder	BMI Category	0.535	0.922	0.971	-0.626	0.200	1.596

Fig. 11: Association Rules Results

The results of association rules on this dataset provide accuracy where the data has several different results, the predicted results from Table 3. One of the association rules found through the FP-Growth algorithm shows a support value of 0.535, which means this pattern appears in 53.5% of all data. The confidence value of 0.913 indicates that when the conditions in the rule are met, the probability of the consequent occurring reaches 91.3%, which indicates a high level of confidence.

Table 3: Visualisation Association Results

Metric	Value
Support	0.535
Confidence	0.913
LaPlace	0.968
Gain	-0.636
P-S	0.197
Lift	1.581

The lift of 1.581 indicates that the relationship between the antecedent and consequent is stronger than a random relationship; that is, the occurrence of the consequent is 1.58 times more likely to occur when the antecedent is met. The LaPlace value of 0.968 also supports a high and stable probability. Although the Gain value is negative (-0.636), this needs to be analyzed further because it could be related to the complexity or limited new information from the rule. The P-S value of 0.197 indicates that the potential for new information from this rule is moderate.

Overall, these rules are considered significant and relevant to interpret because they show patterns that frequently appear, have strong correlations, and can be used as a basis for data-based decision making,

especially in the context of Generation Z's health and lifestyle.

Discussion Classification (Decision Tree)

Subairi *et al.* (2022) used the Gini index and entropy methods for sleep apnea. The results were divided into two categories, namely true and false with a root configuration of 3, shown in Table 4. The table shows that the majority of individuals with poor sleep patterns had high stress levels and daily steps below 4000. This suggests a relationship between sedentary lifestyle and sleep health. However, not all variables had a uniform impact; for example, heart rate did not show a strong correlation—likely due to the influence of other variables such as caffeine consumption that were not captured in the dataset. The class distribution table shows that the data is dominated by the “Normal” category with a total of 72 entries (from the sum of 12+14+26+11+9), while the “Apnea” category only appears in 6 entries. This imbalance reflects the presence of class imbalance in the dataset, which can affect the performance of the classification model—especially in detecting apnea cases. The model tends to be more accurate in predicting the majority category (Normal), but risks ignoring or incorrectly predicting the minority category (Apnea), which is important in the context of health applications. Therefore, techniques such as resampling (oversampling/undersampling) or adjusting class weights are recommended to increase sensitivity to Apnea cases.

Table 4: Discussion Decision Tree

Class	Value
Normal	12
Normal	14
Normal	26
Normal	11
Apnea	1
Normal	9
Apnea	5

Subairi *et al.* (2022) uses the Gini index method as an algorithm to provide decision tree results, while in this study with the Sleep Health and Quality of Life dataset, we use the general decision tree method to produce a performance vector that provides accurate predictions.

Discussion Association (FP Growth)

Perez-Pozuelo *et al.* (2020) using association rules to find the average MCC, accuracy, and Macro F1 with Cross-validated Training set to compare the basis that is always selected by the class that has frequency. By implementing the algorithm that has been compiled, it will produce a matrix based on the evaluation of metrics that have been compiled by the software.

The advancement of sleep digitalization and widespread monitoring technologies is expected to significantly impact how sleep is characterized, diagnosed, and treated. Research by Seow *et al.* (2020)

identified the strongest link between sleep quality and mental health, while the associations with physical and cognitive health were moderate and generally consistent throughout adulthood. In older adults, persistent insomnia symptoms were linked to a notable decline in both mental health (difference = -6.9; SE = 0.4) and physical health (difference = -2.8; SE = 0.4). Additionally, repeated occurrences of both long and short sleep durations were primarily associated with reduced physical well-being (difference = -3.5; SE = 0.9).

The Decision Tree model was evaluated using accuracy, precision, recall, and F1-score metrics. Accuracy reached 93.94%, but the prediction for insomnia was only 56.92% indicating class imbalance. A comparison will be conducted with Random Forest and XGBoost algorithms in further research to evaluate more stable performance and resistance to overfitting.

While Decision Tree and FP-Growth provided interpretable and meaningful results, future work will benchmark these results against ensemble methods such as Random Forest or XGBoost to evaluate model stability and accuracy improvements, particularly in imbalanced datasets.

Based on the patterns found, it is recommended that digital health platforms targeting Gen Z consider daily step monitoring and stress reporting features, as well as providing AI-based adaptive feedback. Educational institutions can also leverage these insights to design data-driven interventions such as healthy sleep and daily physical activity campaigns.

While this study does not propose a new algorithm, it provides an integrated analysis combining classification and association rule mining to explore health patterns among Generation Z — a demographic rarely explored in prior studies using data mining techniques.

Conclusion

The analysis of sleep health and lifestyle data using RapidMiner has provided valuable insights into the classification and association patterns within the dataset. The classification method, employing decision trees, successfully identified key predictors and their associated accuracy. The decision tree visualization revealed significant attribute values such as Blood Pressure Low (90.500), Heart Rate (83), Blood Pressure High (131), Quality of Sleep (5.500), Blood Pressure Low (87.500), BMI Category (split value), and Stress Level (6.500). These findings highlight the most influential factors in understanding sleep health and lifestyle trends among Generation Z.

Additionally, the association rule analysis uncovered relationships between variables using evaluation metrics such as Support (0.535), Confidence (0.913), Laplace (0.968), Gain (-0.636), P-S (0.197), and Lift (1.581). These metrics provided insights into the frequency of co-

occurring itemsets, the reliability of the derived rules, and the likelihood of specific outcomes based on known conditions.

Support refers to how frequently a rule appears in the dataset. A support of 0.535 means the rule applies to 53.5% of the data. Confidence measures the probability of the consequent given the antecedent; 0.913 means there's a 91.3% chance the outcome will occur when the conditions are met. Lift measures the rule's strength compared to random chance; a lift of 1.581 indicates the rule is 1.58 times more likely to occur than by chance.

The combined use of classification and association rule analysis enabled a comprehensive understanding of the interplay between sleep health and lifestyle attributes. The results can inform targeted interventions, personalized recommendations, and evidence-based strategies to improve health outcomes for Generation Z. This study emphasizes the potential of data mining techniques to uncover meaningful patterns and support data-driven decision-making in the domain of public health.

The patterns identified in this study offer a foundation for digital health interventions aimed at Generation Z. Applications such as wearable-based step trackers with personalized alerts or AI-based stress monitoring systems could be developed. Moreover, educational institutions and health policymakers should leverage these findings to craft preventive programs targeting sedentary behavior and poor sleep hygiene, aligning with digital health transformation agendas.

Acknowledgment

The authors would like to express their sincere gratitude to the University of Bunda Mulia, particularly the Center for Research and Community Service (Pusat Penelitian dan Pengabdian Masyarakat – P3M), for their valuable support in making this research possible.

Funding Information

The researchers would like to thank to University of Bunda Mulia who has funded this research so that it can run well and as expected.

Authors Contributions

Francka Sakti Lee: Conceptualization, methodology design, and data analysis.

Johanes Fernandes Andry: Data collection and preprocessing.

Kevin Christianto: Algorithm development and implementation.

Yunianto Purnomo: Statistical analysis and result interpretation.

Aziza Chakir: Supervision, validation, and review of the manuscript.

Lina Noviana: Literature review and manuscript drafting.

Ethics

This article does not deal with human or animal subjects. In addition, this article is original and has never been published. The corresponding author confirms that all other authors have read and agree that the manuscript does not involve ethical issues.

References

- Andry, J. F., Reynaldo, S. A., Christianto, K., Lee, F. S., Loisa, J., & Manduro, A. B. (2021). Algorithm of Trending Videos on YouTube Analysis using Classification, Association and Clustering. *2021 International Conference on Data and Software Engineering (ICoDSE)*, 1–6.
<https://doi.org/10.1109/icodse53690.2021.9648486>
- Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of Big Data*, 9(1), 24. <https://doi.org/10.1186/s40537-021-00553-4>
- Dinata, D. F., Lee, F. S., Geasela, Y. M., Everlin, S., & Purnomo, Y. (2024). Website-Based Educational Application to Help MSMEs in Indonesia Develop. *Journal of Computer Science*, 20(7), 742–750.
<https://doi.org/10.3844/jcssp.2024.742.750>
- Francis, B. K., & Babu, S. S. (2019). Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems*, 43(6), 204.
<https://doi.org/10.1007/s10916-019-1295-4>
- Geasela, Y. M., Bernanda, D. Y., Andry, J. F., Jusuf, C. K., Winata, S., Lydia, & Everlin, S. (2024). Analysis of Student Mental Health Dataset Using Mining Techniques. *Journal of Computer Science*, 20(1), 121–128.
<https://doi.org/10.3844/jcssp.2024.121.128>
- Gupta, M. K., & Chandra, P. (2020). A Comprehensive Survey of Data Mining. *International Journal of Information Technology*, 12(4), 1243–1257.
<https://doi.org/10.1007/s41870-020-00427-7>
- Hale, L., Troxel, W., & Buysse, D. J. (2020). Sleep Health: An Opportunity for Public Health to Address Health Equity. *Annual Review of Public Health*, 41(1), 81–99.
<https://doi.org/10.1146/annurev-publhealth-040119-094412>
- Haryana, N. R., Rosmiati, R., Purba, E. M., & Firmansyah, H. (2023). Gaya Hidup Generasi Z Dalam Konteks Perilaku Makan, Tingkat Stres, Kualitas Tidur dan Kaitannya Dengan Status Gizi: Literature Review. *Jurnal Gizi Kerja Dan Produktivitas*, 4(2), 267.
<https://doi.org/10.62870/jgkp.v4i2.24990>
- Javaid, M., & Haleem, A. (2019). Industry 4.0 applications in medical field: A brief review. *Current Medicine Research and Practice*, 9(3), 102–109.
<https://doi.org/10.1016/j.cmrp.2019.04.001>
- Madyatmadja, E. D., Rianto, A., Andry, J. F., Tannady, H., & Chakir, A. (2021). Analysis of Big Data in Healthcare Using Decision Tree Algorithm. In *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)* (pp. 111–115).
<https://doi.org/10.1109/iccsai53272.2021.9609734>
- Omoyiola, B. O. (2023). The social implications, risks, challenges and opportunities of big data. *Emerald Open Research*, 1(4), 46554.
<https://doi.org/10.1108/eor-04-2023-0014>
- Perez-Pozuelo, I., Zhai, B., Palotti, J., Mall, R., Aupetit, M., Garcia-Gomez, J. M., Taheri, S., Guan, Y., & Fernandez-Luque, L. (2020). The future of sleep health: a data-driven revolution in sleep science and medicine. *Npj Digital Medicine*, 3(1), 42.
<https://doi.org/10.1038/s41746-020-0244-4>
- Ruan, Q., Yang, K., Wang, W., Jiang, L., & Song, J. (2020). Correction to: Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Medicine*, 46(6), 1294–1297.
<https://doi.org/10.1007/s00134-020-06028-z>
- Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., & Shapi'i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in Biology and Medicine*, 136, 104754.
<https://doi.org/10.1016/j.compbiomed.2021.104754>
- Seow, L. S. E., Tan, X. W., Chong, S. A., Vaingankar, J. A., Abdin, E., Shafie, S., Chua, B. Y., Heng, D., & Subramaniam, M. (2020). Independent and combined associations of sleep duration and sleep quality with common physical and mental disorders: Results from a multi-ethnic population-based study. *PLOS ONE*, 15(7), e0235816.
<https://doi.org/10.1371/journal.pone.0235816>
- Sissoko, O. A. A., & Prasetyawati, H. (2022). Kebutuhan Gaya Hidup Generasi Z Terhadap Perilaku Narsis di Instagram. *Matriks Jurnal Sosial Dan Sains*, 4(1), 31–40.
<https://doi.org/10.59784/matriks.v4i1.128>
- Subairi, S., Permatasari, D. C., Dirgantara, W., Surya Akbar Gumilang, Y., Zahroya J.M.F., I., & Haitsam, H. (2022). Deteksi Sleep Apnea Menggunakan Metode Decision Tree dengan Fitur Statistik RR Interval. *Jurnal EECCIS (Electrics, Electronics, Communications, Controls, Informatics, Systems)*, 16(3), 96–100.
<https://doi.org/10.21776/jeeccis.v16i3.1603>
- Tharmalingam, L. (2023). *Sleep Health and Lifestyle Dataset*.
<https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

- Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Dellinger, F. N., Djousse, L., Elkind, M. S. V., Ferguson, J. F., Fornage, M., Khan, S. S., Kissela, B. M., Knutson, K. L., Kwan, T. W., Lackland, D. T., ... Tsao, C. W. (2020). Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association. *Circulation*, 141(9), e139–e596. <https://doi.org/10.1161/cir.0000000000000757>
- Wijoyo, H., Indrawan, I., Cahyono, Y., Handoko, A. L., & Santamoko, R. (2020). *Generasi Z Dan Revolusi Industri 4.0*.
- Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Information Fusion*, 77, 29–52. <https://doi.org/10.1016/j.inffus.2021.07.016>