ComTech: Computer, Mathematics and Engineering Applications, 16(1), June 2025, 35−49 P-ISSN: 2087-1244
DOI: 10.21512/comtech.v16i1.12265 E-ISSN: 2476-907X

Implementation of Decision Tree with Best Subset Approach to Identify Suicide Cases in Central and East Java

Moh. Miftahul Choiri[1], Denny Nurdiansyah[2]*, Auliyaur Rokhim[3], Pebriana Putri Patmawati[4], and Putri Vatria Pebriani[5]

[1-5]Statistics Study Program, Faculty of Science and Technology, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Indonesia 62115
[1]choiri01@unugiri.ac.id; [2]denny.nur@unugiri.ac.id;
[3]aurrokhim@unugiri.ac.id; [4]pebrianaputri65@gmail.com; [5]febrianivatria@gmail.com

Received: 27th September 2024/ **Revised:** 2nd December 2024/ **Accepted:** 3rd December 2024

How to Cite: Choiri, M. M., Nurdiansyah, D., Rokhim, A., Patmawati, P. P., & Pebriani, P. V. (2025). Implementation of decision tree with best subset approach to identify suicide cases in Central and East Java. *ComTech: Computer, Mathematics and Engineering Applications, 16*(1), 35−49. https://doi.org/10.21512/comtech.v16i1.12265

Abstract **-** The research was conducted to determine the descriptive statistics of suicide cases and classify suicide cases based on the attributes of victims who made suicide attempts. The research design used was a quantitative method in the form of exploratory research using the Decision Tree method. The research novelty was applying the Decision Tree method with the Best Subset approach. The research data sources were obtained from online mass media news such as DetikJatim and DetikJateng for suicide attempt cases from January 2022 to July 2024. The research finds significant differences in the number of suicide attempts in East Java and Central Java, with Surabaya, Malang, Blitar, Semarang, and Klaten recording higher numbers. The findings show that males more often attempt suicide, while females more often experience failed attempts. Young adults (20−39 years) record the highest rate, and hanging is the most common method. Unknown mental disorders and depression are the main risk factors, with many attempts occurring without rescue. The implication is that improving emergency response systems and mental health services is essential. The research recommends strengthening mental health and social support for older adults and those under stress. Then, enhancing rapid rescue efforts with comprehensive psychological interventions is essential for suicide prevention. The originality of the research lies in the use of a Decision Tree with the Best Subset approach to identify suicide patterns based on risk factors and methods used.

Keywords**:** Decision Tree method, Best Subset approach, suicide cases

I. INTRODUCTION

Suicide is the act of someone deliberately ending his/her life (Jakaria et al., 2023). This severe act involves a variety of influential factors and complexities, such as mental health issues, emotional distress, and environmental factors. If someone is experiencing suicidal feelings, it is imperative to seek immediate help from a mental health professional or health institution that can provide the necessary support and intervention. According to the World Health Organization (WHO), more than 800,000 people commit suicide each year worldwide, and suicide accounts for more than half of the world's violent deaths (Wasserman et al., 2020). One of the causes of suicide is bipolar disorder in individuals due to risk factors such as individuals with male gender, living alone, divorced, childless, certain races, younger age (< 35 years), older age (> 75 years), not working, and personal history of suicide attempts and family history of suicide attempts or suicide, as well as dominant depressive polarity (Miller & Black, 2020).

The COVID-19 pandemic has had a vast mental health impact around the world. It is expected that suicide rates will increase due to the long-term social, economic, and mental health impacts on the global community. There is evidence of suicide deaths due to a pandemic, such as an increase in suicides in the United States during the Influenza pandemic of 1918−1992 and suicides by older adults in Hong Kong during the Severe Acute Respiratory Syndrome (SARS) epidemic of 2003 (Gunnell et al., 2020). In Indonesia, several variables affect the mental health of people with COVID-19, namely anxiety, suicide, and

*Corresponding Author

insomnia (Wahyuningrum et al., 2022). The world of education at that time is also affected by the existence of online teaching, which makes someone lonely and raises suicidal ideas (Hamzah & Triwahyuni, 2023). However, online learning in certain areas is still acceptable (Nurdiansyah, 2023). Other contributing factors to suicide during the COVID-19 pandemic are psychological and economic factors due to mass layoffs during the pandemic (Sianturi & Zulaeha, 2022). Until now, suicide remains a crucial mental problem, like a couple who committed suicide due to debt problems (Krista, 2024), a student who committed suicide due to depression due to an unfinished thesis (Aminudin, 2024), and three incidents of family suicide in East Java (Baihaqi, 2023). Based on police data up to May 24, 2023, there have been 451 cases of suicide in Indonesia. The largest number is in Central Java, followed by East Java, Bali, West Java, Yogyakarta, and North Sumatra (Purbaya, 2023).

The research topic raised in this study is the application of classification methods for suicide attempt cases. Common problems that occur, usually suicide attempts, receive less attention from local governments and surrounding communities that are close to suicide victims. In addition, an input of knowledge related to the rules or mindset of suicide victims is needed to assist community service activities that require material expertise for more effective and efficient socialization or counselling related to the prevention of suicide attempts. The problem of suicide can be prevented by identifying the possibility of a suicide attempt based on the attributes of the attempted victim that support individuals to have the desire to commit suicide. For this problem, it is necessary to identify the possibility of suicide attempts based on individual status, environmental conditions, and triggering factors as a precursor to preventing suicide attempts. The solution to this problem is to obtain rules from the suicide attempt pattern, which is the result of knowledge from classification methods such as the decision tree method.

The research purpose is to determine the descriptive statistics of suicide cases and classify suicide cases based on the attributes of victims who make suicide attempts. The research method used is quantitative in the form of exploratory research using the Decision Tree method. This method is prevalent in classification studies in the scientific field of data mining.

The Decision Tree method is the most popular method for applying data classification studies, which is helpful in the scientific fields of statistics, pattern recognition, and machine learning (Jijo & Abdulazeez, 2021). Decision Trees are part of predictive analysis in business decision-making in various fields (Lee et al., 2022). Decision Tree is also a machine learning algorithm classified as supervised learning, which in previous studies was applied by comparing the separation index, namely the Gini index and Information Gain with the classification accuracy of the model estimated using different matrices such as confusion matrix, overall accuracy, per-class accuracy, recall, and precision (Tangirala, 2020). Decision Tree improves its performance by accelerating the processing time with the Boosting algorithm (Zhao et al., 2021). In previous research, a Decision Tree is applied based on Ensemble Learning by implementing the Random Forest and Extremely Randomized Trees or Extra Trees algorithms (Ghiasi & Zendehboudi, 2021). However, some of these methods are concluded only to organize the dataset and attribute selection randomly. It needs to be optimal because the dataset formed does not necessarily result in optimal Decision Tree performance.

Research on subsets in Decision Tree models has evolved significantly over time. Previous research has laid an early foundation by comparing heuristic and complete search strategies for feature subset selection in decision trees, addressing the inefficiencies of greedy methods, and advocating for alternatives like branch-and-bound to enhance subset quality (Murthy, 1998). Building on this, a caching mechanism is introduced to optimize feature subset exploration in Decision Tree learning, reducing redundant computations and improving efficiency while maintaining accuracy (Caruana & Freitag, 1994).

Later, a groundbreaking approach is proposed by framing decision tree construction as a mixed-integer optimization problem. The method simultaneously optimizes tree structure and feature splits, focusing on balancing error minimization and tree simplicity for scalable applications (Bertsimas & Dunn, 2017). In a significant contribution, another research by Ruggieri (2019) has developed a complete search algorithm to identify optimal feature subsets in Decision Trees. It emphasizes exact enumeration and pruning techniques to avoid overfitting, offering a robust solution for feature selection in large datasets. Previous researchers have also refined Decision Tree optimization further through the integration of Stochastic Gradient Boosting (SGB) loss functions, as reported in SpringerOpen Journal. This method significantly enhances predictive performance by addressing overfitting and improving convergence rates. It is particularly impactful in domains like medical diagnostics, where precision and robustness are critical (Jawalkar et al., 2023). Finally, a hybrid methodology combining genetic algorithms and decision trees is introduced (Gkikas et al., 2023). This approach identifies optimal subsets by minimizing overfitting and showcases its utility in real-world scenarios, such as wine classification, further advancing the field. These previous studies collectively highlight the progression from foundational theories to advanced algorithms for optimizing feature subsets in Decision Tree models.

The novelty of this research is that it will apply the Decision Tree with the Best Subset approach. This approach leads to setting all the possibilities of the dataset attributes directed to the label attribute. This approach mimics the Best Subset method in discussing regression analysis related to selecting

predictor variables by evaluating standard deviation and R-Square (Draper & Smith, 1998). In addition, the novelty offered in the research is the application of the Decision Tree on limited suicide case data from observing news circulating in the online mass media. The research contributes to identifying patterns and risk factors for suicide using an innovative approach with the Decision Tree method and Best Subset analysis, offering new insights into data analysis.

The research has significant benefits in providing a deeper understanding of the patterns and factors that influence suicide attempts, mainly through descriptive statistical analysis and classification based on victim attributes. By using the Decision Tree method, which is popular in classification studies in the field of Data Mining, the researchers hope that the research can identify vulnerable groups as well as key risk factors that contribute to suicide attempts. The urgency of the research lies in its ability to support more effective prevention efforts and targeted interventions, thereby helping to reduce the number of suicide attempts through data-driven and more targeted mental health policies and programs.

II. METHODS

The research design used is a quantitative method in the form of exploratory research using the Decision Tree method. The research novelty is applying the Decision Tree method with the Best Subset approach. This method is applied by creating algorithms and programming with Open-Source R software. R software is a statistical software that is license-free and easy to get on the official website (Nurdiansyah & Sulistiawan, 2023).

In the development of the use of R software in data mining, program packages are available for applying Decision Tree methods, such as gtools, rpart, caret, e1071, and rpart.plot packages. The gtools package is used to generate combinations and permutations (Warnes et al., 2023). Then, rpart builds Decision Tree models for classification and regression (Therneau & Atkinson, 2025). At the same time, caret provides a framework for model training and evaluation, including an accuracy matrix (Kuhn, 2024). The e1071 package supports caret by providing a confusion matrix function to assess model performance (Meyer et al., 2024), and rpart.plot allows visualization of the Decision Tree generated by rpart, making it easy to interpret model results graphically (Milborrow, 2024). The combination of these packages supports comprehensive data mining analysis.

These program packages need to be installed online on the laptop because they support the process of making algorithms and programming the R language using R software to develop the Best Subset-based Decision Tree program. The research population is determined from all attempted suicide cases in Indonesia, specifically in East Java and Central Java. Samples are selected from news in online mass media for suicide attempt cases. The research location is at the Research Laboratory of the Statistics Study Program, Nahdlatul Ulama Sunan Giri University (UNUGIRI).

Purposive sampling is used in the research. Historical data samples of news recorded in online mass media on the DetikJatim and DetikJateng web pages are used. The research data source is obtained from online mass media news for suicide attempt cases from January 2022 to July 2024. In defining the research variables, the label attribute is given to the suicide attempt variable (Y). In contrast, the ordinary attributes are represented by gender (X_1), age (X_2), rescue action (X_3), attempt method (X_4), mental disorder (X_5), and trigger risk (X_6). The Y and X_3 have a binomial data type, while the other variables are given a polynomial data type. The research variables are described in Table 1 (see Appendices).

The research steps include several main stages. First, the data set is determined by entering and pre-processing the data. Next, the descriptive statistical results of the suicide attempt cases are displayed and explained. The data mining method used is the Decision Tree classification method with the Best Subset approach. Evaluation of the method is done through training and testing on the full dataset, with performance measures such as the confusion matrix, including accuracy, Kappa, and F1-score. Finally, the research produces findings in the form of knowledge extracted from the Decision Tree, in the form of rules or patterns relevant to suicide cases.

The general Decision Tree procedure starts with preparing the training data. Attributes are selected as roots based on information gain, where the attribute with the largest gain is selected from all existing attributes. The calculation starts by calculating Entropy using Equation (1) to get the Gain value. It has S as the set of cases, n as the number of partitions of S, and p_i as the proportion of S_i to S. The index i runs from 1 to n to calculate the contribution of each category to the total Entropy value. After that, the Gain value is calculated by Equation (2). The A is the attribute, $|S_i|$ is the number of cases in the i-th partition, and $|S|$ is the total number of cases in S. Next, branches are created for each attribute value, and this process is repeated for each branch until all cases in a branch have the same class. Here, the index i denotes each partition of the data based on attribute A.

$$Entropy(S) = -\sum_{i=1}^{n}(p_i * {}^{2}log\ p_i) \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n}\left(\frac{|S_i|}{|S|} * Entropy(S_i)\right) \quad (2)$$

The Decision Tree procedure with the Best Subset approach starts with entering the observation data and determining the data type of each attribute. Then, all possible combinations of subsets of different attributes are determined, so various datasets are formed based on the existing combinations. Each dataset formed is evaluated by the classification method, and

its performance is measured using accuracy, Kappa, and F1-score. After that, the method results for each dataset are compared to find the combination with the highest accuracy, Kappa, and F1-score values. Finally, the visualization of the Decision Tree and the rules formed from the best method are shown and explained.

III. RESULTS AND DISCUSSIONS

Observations are made to examine cases of attempted suicide in East Java and Central Java. Suicide attempt data are collected from 223 news articles in East Java and 84 news articles in Central Java, observed from January 2022 to July 2024 on the DetikJatim and DetikJateng web pages. The 307 observation data collected reveals that Surabaya records the highest number of suicide attempts, with 40 cases. The result is followed by Malang with 26 attempts, Blitar with 24 attempts, Semarang with 19 attempts, and Klaten with 13 attempts, as shown in Figure 1 (see Appendices).

From Figure 1, there is a significant difference in the number of suicide attempts across different cities in East Java and Central Java, with Surabaya, Malang, Blitar, Semarang, and Klaten recording higher rates of suicide attempts than other regions. These high cases suggest the presence of unique factors that need to be further investigated. Prevention approaches tailored to the local context, improved access to mental health services, as well as public awareness campaigns, are essential to address this issue effectively.

Observations of 307 suicide attempt news stories reveal that the highest number of suicide attempts are made by men, with 183 suicide attempts and 32 failed attempts. Women come second with 61 suicide attempts and 31 failed attempts. The results are shown in Figure 2 (see Appendices). From Figure 2, it is found that men attempt suicide more often than women, which may be due to men's tendency to seek less emotional support. While women make fewer suicide attempts, they show a high rate of failed attempts, which can be due to the use of less lethal methods or other factors that prevent success. These findings emphasize the importance of a prevention approach that considers gender differences in suicide treatment and prevention efforts.

The analysis of 307 suicide attempt news by age shows that the highest number of suicide attempts occurs in the young adult group, with 97 attempts. The adult group in second place records 79 suicide attempts, while the adolescent group records 27 attempts, and the elderly group has 22 attempts. In addition, there are 22 failed suicide attempts in the young adult group, followed by adolescents with 18 failed attempts, as shown in Figure 3 (see Appendices).

From Figure 3, it is found that the young adult group (20-39 years) has the highest number of suicide attempts, probably due to the significant life stress at this age. The mature group (40-59 years) also shows high rates, while adolescents and the elderly have lower but still significant rates. The high rate of failed attempts among young adults and adolescents signals the need to focus on appropriate mental health interventions and support to address the specific challenges faced by each age group.

An examination of 307 suicide attempt news based on rescue actions shows that suicide attempts without rescue reach the highest number of 214 attempts. Unsuccessful suicide attempts with rescue take second place with 59 attempts. In third place, there are 30 suicide attempts despite rescue, as shown in Figure 4 (see Appendices).

From Figure 4, most suicide attempts occur without rescue, indicating the need for improvement in the emergency response system. Although there are a few cases of attempts that are successfully rescued or failed with rescue, the number is still relatively low. This finding emphasizes the importance of improving the rescue and preparedness system to deal with suicide attempts more effectively.

An examination of 307 news of suicide attempts based on the method used reveals that hanging (Method1) ranks highest with 129 suicide attempts. In second place, taking drugs, poisons, or dangerous chemicals (Method2) records 27 suicide attempts. The third place is jumping into a river, lake, pond, sea, or well (Method6) with 25 suicide attempts. In fourth place, jumping from a bridge, building, or tall building (Method5) records 24 suicide attempts. In fifth place, running into a train (Method7) has 22 suicide attempts. In addition, Method5 and Method6 show 24 and 11 failed attempts, respectively, as shown in Figure 5 (see Appendices).

From Figure 5, it is found that hanging is the most common method of attempted suicide, followed by drinking chemicals, jumping into water, and jumping from a tall place. The high number of failed attempts for jumping from a tall place and into water indicates that not all attempts with these methods are fatal. These findings emphasize the importance of method-specific treatment and the need for effective prevention strategies for each method to reduce risk and improve intervention outcomes.

An analysis of 307 suicide attempt news by mental disorder shows that suicide attempts with an unknown mental disorder (Disorder0) have the highest number of 165 attempts. Suicide attempts with depression (Disorder1) come second with 71 attempts. In addition, failed suicide attempts with Disorder0 and Disorder1 record 49 and 14 attempts, respectively, as shown in Figure 6 (see Appendices).

From Figure 6, it is found that an unknown mental disorder is the most common factor in suicide attempts, followed by depression. The high rate of failed attempts among individuals with unknown mental disorders and depression suggests the need for improvements in mental health assessment and support. The result underscores the importance of early detection and better interventions to address unidentified mental disorders, as well as providing more effective support for those with depression.

Then, 307 suicide attempt news stories by trigger risk show that unknown trigger (Risk0) is the highest trigger, with 120 suicide attempts, with 33 failed attempts. The second trigger is debt, online loans, or financial problems (Risk1) with 28 attempts. The third trigger is prolonged illness (Risk6) with 21 attempts. The fourth trigger is breakup, cheating victim, or romance problem (Risk2). It has 20 attempts with 13 failed attempts, as shown in Figure 7 (see Appendices).

From Figure 7, unknown triggers are the main factor in suicide attempts, followed by financial problems and prolonged illness. Romance problems also contribute significantly, with several failed attempts. These findings emphasize the need for a better understanding of specific triggers as well as more targeted prevention strategies to address the various triggering factors that can lead to suicide attempts.

In the application of the Decision Tree method with the Best Subset approach, the results of the possible subsets are formed, and their performance values in R software are shown in Table 2 (see Appendices). In Table 2, The best subset selection results show that for two-variable subsets, the combination of X_3 and X_6 yields the best performance with an accuracy of 0.9153, a Kappa of 0.7601, and an F1-Score of 0.8143. For three-variable subsets, the best subsets are X_3, X_4, and X_6 with an accuracy of 0.9283, a Kappa of 0.7829, and an F1-Score of 0.8281. For four-variable subsets, the two best subsets achieve the same performance with an accuracy of 0.9283, a Kappa of 0.7854, and an F1-Score of 0.8308, and this result is consistent for five-variable subsets as well. The X_5 appears to contribute less, as subsets include lower performance or missing metric values. Therefore, selecting the right combination of variables is crucial to obtain an optimal predictive model without including less relevant variables. The best results from each subset in Table 2 are highlighted with yellow shading as markers and compiled in Table 3 (see Appendices) for comparison.

In Table 3, among the eight subset options, several subsets share the same highest performance values, except for the first and second subset options. The application of the Decision Tree using subset options three through eight resulted in decision rules derived from two types of subsets: the first subset ($X_2X_3X_4X_6$) and the second subset ($X_3X_4X_5X_6$). The decision rules from the first subset are presented in Figure 8 (see Appendices), while those from the second subset are shown in Figure 9 (see Appendices).

In Figure 8, Decision Tree is obtained with knowledge in the form of rules that can be translated into suppositions. First, suicide will occur if there is no rescue action. Second, suicide will also occur if there are triggering risks such as Risk4 (killing spouse or family), Risk5 (becoming a criminal suspect), Risk6 (prolonged illness), or Risk7 (following family death), even if there is a rescue action. Third, the suicide event will fail if the triggering risks are Risk0 (unknown), Risk1 (debt, online loan, or financial problems), Risk2 (breaking up, victim of cheating, or romantic problems), Risk3 (divorce), Risk8 (loss of job (layoff)), Risk11 (conflict with family, spouse, or children), or Risk12 (other triggers); the method of attempt is Method2 (taking drugs, poisons, or dangerous chemicals), Method3 (cutting the neck or abdomen with a knife), Method5 (jumping from a bridge or building), Method7 (crashing into a train), or Method8 (other methods); and there is a rescue action. Fourth, suicide will also fail if the method used is Method1 (hanging oneself), Method4 (burning oneself), or Method6 (jumping into a river, lake, pond, sea, or well), a trigger risk such as Risk1, Risk2, or Risk11, and a rescue action. Fifth, suicide will fail if the attempted method is Method1, Method4, or Method6; there are risks such as Risk0 or Risk8, the victim is categorized as Adult, Juvenile, or unknown, and there is a rescue action. Finally, suicide will occur if the method of attempt is Method1, Method4, or Method6; there are risks such as Risk0 or Risk8, and the victim is categorized as elderly, child, or young adult, although there is a rescue.

Furthermore, in Figure 9, a Decision Tree is obtained with a rule that becomes the second decision option that can be translated into supposition language. First, suicide will occur if there is no rescue action. Second, suicide will also occur if there are triggering risks such as Risk4 (killing a spouse or family), Risk5 (becoming a criminal suspect), Risk6 (prolonged illness), or Risk7 (following a family death), even if there is a rescue action. Third, suicide will fail if the triggering risks are Risk0 (unknown), Risk1 (debt, online loan, or financial problems), Risk2 (breakup, victim of cheating, or romantic problems), Risk3 (divorce), Risk8 (loss of job (layoff)), Risk11 (conflict with family, spouse, or children), or Risk12 (other triggers); the attempted method is Method2 (taking drugs, poisons, or dangerous chemicals), Method3 (cutting the neck or abdomen with a knife), Method5 (jumping from a bridge or building), Method7 (crashing into a train), or Method8 (other methods); and there is a rescue action. Fourth, suicide will fail if the attempted method used is Method1 (hanging oneself), Method4 (burning oneself), or Method6 (jumping into a river, lake, pond, sea, or well), with a trigger risk of Risk1, Risk2, or Risk11, and there is a rescue action. Fifth, suicide will also fail if the method is Method1, Method4, or Method6; there are risks such as Risk0 or Risk8, the victim has a mental disorder of Disorder0 (unknown); and there is a rescue action. Finally, suicide will occur if the attempted method is Method1, Method4, or Method6, there is a risk of Risk0 or Risk8, and the victim has a mental disorder of Disorder1 (depression), although there is a rescue action.

These two graphs use a Decision Tree to identify patterns of suicide incidence based on the risk factors and experimental methods used. In Figure 8, the decision rules show that suicide is likely to occur in the absence of rescue attempts or when triggering risks such as killing family, being a criminal suspect,

prolonged illness, or following a family death occur, despite rescue attempts. Suicidal events can be prevented if there are rescue efforts, attempted methods, or certain risk factors such as financial problems, romance, and family conflict. In addition, the age factor also influences the outcome, where the elderly or young adults are more prone to failing a rescue. In Figure 9, a similar pattern is found, but with the addition of the mental disorder factor, where victims with depression have a higher risk of suicide, even in the presence of rescue measures.

The two graphs reveal that the risk of attempted suicide is strongly influenced by a combination of triggering factors, method of attempt, age of victim, and presence of mental disorders. Rescue attempts can be effective in some situations, but certain risks, such as depression or severe social factors, increase the chances of suicide even with intervention. The method of attempt and age are also important factors in determining the success of the rescue.

IV. CONCLUSIONS

The research analyzes descriptive statistics and classifies suicide attempt cases based on victim attributes using the Decision Tree method with the Best Subset approach. The results reveal significant regional differences, with higher rates of suicide attempts recorded in Surabaya, Malang, Blitar, Semarang, and Klaten. Males are more likely to attempt suicide, while females experience a higher rate of failed attempts. The most vulnerable age group is 20-39 years, with hanging being the most common method. Key risk factors include depression and undiagnosed mental disorders, with many cases occurring without rescue. The research highlights the need to enhance mental health services, emergency response systems, and comprehensive psychological interventions for suicide prevention.

The research highlights the need for enhanced mental health support systems, targets interventions for high-risk groups like young adults, and improves emergency response efforts to prevent suicide attempts. The use of the Decision Tree method with the Best Subset approach offers a novel way to identify risk patterns and inform prevention strategies. However, the research is limited by its reliance on online mass media data, which may omit unreported cases, and its focus on East Java and Central Java, restricting generalizability. Additionally, the lack of variables such as socioeconomic status or access to healthcare limits the scope of analysis. Future research should address these gaps using more comprehensive datasets and broader variables such as education level, income, employment status, family background, and access to mental health services. It can provide deeper insights into the underlying causes of suicide attempts and support more effective, targeted prevention policies.

ACKNOWLEDGEMENT

The research was funded by a grant from Universitas Nahdlatul Ulama Sunan Giri (UNUGIRI) in 2024. The authors would like to express their gratitude to the Faculty of Science and Technology for providing additional financial support to facilitate this research. The support from the Research and Community Engagement Office is also deeply appreciated.

The authors also acknowledge the financial support from Directorate General of Higher Education (Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM)), Ministry of Education, Culture, Research, and Technology (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia (Kemendikbudristek)). This funding, provided under the 2024 fiscal year research program, played a significant role in the successful completion of the research.

AUTHOR CONTRIBUTIONS

Conceived and designed the analysis, M. M. C. and A. R.; Collected the data, P. P. P. and P. V. P.; Contributed data or analysis tools, D. N., P. P. P., and P. V. P.; Performed the analysis, D. N.; and Wrote the paper M. M. C. and A. R.

DATA AVAILABILITY

The data that support the findings of the research are openly available in Zenodo at https://doi.org/10.5281/zenodo.15486188, reference number: 15486188. The dataset is compiled by the authors based on publicly accessible online news sources. The data are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

REFERENCES

Aminudin, M. (2024, January 9). *Motif mahasiswa Malang bunuh diri gegara depresi skripsi tak selesai*. Detikjatim. https://www.detik.com/jatim/berita/d-7132997/motif-mahasiswa-malang-bunuh-diri-gegara-depresi-skripsi-tak-selesai

Baihaqi, A. (2023, December 19). *Tiga tragedi satu keluarga bunuh diri yang gemparkan Jatim*. Detikjatim. https://www.detik.com/jatim/berita/d-7097707/tiga-tragedi-satu-keluarga-bunuh-diri-yang-gemparkan-jatim

Bertsimas, D., & Dunn, J. (2017). Optimal classification trees. *Machine Learning*, *106*, 1039–1082. https://doi.org/10.1007/s10994-017-5633-9

Caruana, R., & Freitag, D. (1994). Greedy attribute selection. In *Machine Learning Proceedings 1994* (pp. 28–36). Elsevier. https://doi.org/10.1016/B978-1-55860-335-6.50012-X

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. John Wiley & Sons, Inc.

Ghiasi, M. M., & Zendehboudi, S. (2021). Application of decision tree-based ensemble learning in the classification of breast cancer. *Computers in Biology and Medicine*, *128*. https://doi.org/10.1016/j.compbiomed.2020.104089

Gkikas, D. C., Theodoridis, P. K., Theodoridis, & Gkikas, M. C. (2023). Finding good attribute subsets for improved decision trees using a genetic algorithm wrapper; a supervised learning application in the food business sector for wine type classification. *Informatics*, *10*(3), 1–30. https://doi.org/10.3390/informatics10030063

Gunnell, D., Appleby, L., Arensman, E., Hawton, K., John, A., Kapur, N., ... & Yip, P. S. (2020). Suicide risk and prevention during the COVID-19 pandemic. *The Lancet Psychiatry*, *7*(6), 468–471. https://doi.org/10.1016/S2215-0366(20)30171-1

Hamzah, V. D. S., & Triwahyuni, A. (2023). Kesepian sebagai faktor prediktif ide bunuh diri dikalangan mahasiswa: Studi pada mahasiswa Universitas Padjadjaran selama pandemi COVID-19. *MEDIAPSI*, *9*(2), 85–97. https://doi.org/10.21776/ub.mps.2023.009.02.894

Jakaria, Nasution, H. H., Lubis, F. H., & Ramadi, B. (2023). Analisis hukum Islam tentang bunuh diri dan faktor-faktor pemicunya: Ditinjau dari psikologis, sosial, dan kesehatan mental serta kaitannya dalam pasal 345 KUHP. *Triwikrama: Jurnal Ilmu Sosial, 2*(3), 81–90.

Jawalkar, A. P., Swetcha, P., Manasvi, N., Sreekala, P., Aishwarya, S., Kanaka Durga Bhavani, P., & Anjani, P. (2023). Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting. *Journal of Engineering and Applied Science*, *70*, 1–18. https://doi.org/10.1186/s44147-023-00280-y

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(1), 20–28. https://doi.org/10.38094/jastt20165

Krista, P. (2024, January 12). *Pasutri ketua pecalang bunuh diri di Pantai Gunaksa diduga karena utang*. Detikbali. https://www.detik.com/bali/hukum-dan-kriminal/d-7138841/pasutri-ketua-pecalang-bunuh-diri-di-pantai-gunaksa-diduga-karena-utang

Kuhn, M. (2024, December 10). *caret: Classification and regression training*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=caret

Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive analytics in business analytics: Decision tree. *Advances in Decision Sciences*, *26*(1), 1–29. https://doi.org/10.47654/v26y2022i1p1-30

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2024, September 16). *E1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=e1071

Milborrow, S. (2024, February 26). *rpart.plot: Plot 'rpart' Models: An enhanced version of 'plot.rpart'*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=rpart.plot

Miller, J. N., & Black, D. W. (2020). Bipolar disorder and suicide: A review. *Current Psychiatry Reports*, *22*. https://doi.org/10.1007/s11920-020-1130-0

Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery*, *2*, 345–389. https://doi.org/10.1023/A:1009744630224

Nurdiansyah, D. (2023). The investigation of the acceptance of students against Microsoft Teams learning with the SEM-PLS approach. *Vygotsky: Jurnal Pendidikan Matematika Dan Matematika*, *5*(1), 13–28. https://doi.org/10.30736/voj.v5i1.686

Nurdiansyah, D., & Sulistiawan, A. (2023). *Dasar pemrograman komputer dengan open source software R (Untuk bidang sains dan teknologi)*. CV. AA. Rizky.

Purbaya, A. A. (2023, October 13). *Jateng tertinggi kasus bunuh diri, psikolog: Saatnya menyapa orang terdekat*. Detikjateng. https://www.detik.com/jateng/berita/d-6980270/jateng-tertinggi-kasus-bunuh-diri-psikolog-saatnya-menyapa-orang-terdekat

Ruggieri, S. (2019). Complete search for feature selection in decision trees. *Journal of Machine Learning Research, 20*(104), 1–34.

Sianturi, R., & Zulaeha, A. (2022). Increasing of suicide in Pandemic COVID 19 by depression. *Jurnal Ilmiah Keperawatan Altruistik*, *5*(1), 1–10.

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm*. *International Journal of Advanced Computer Science and Applications (IJACSA), 11*(2), 612–619. https://doi.org/10.14569/IJACSA.2020.0110277

Therneau, T., & Atkinson, B. (2025, January 7). *rpart: Recursive partitioning and regression trees*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=rpart

Wahyuningrum, T., Prasetyo, N. A., Bahtiar, A. R., Latifah, L., Ramadhani, I. D., & Yunitawati, D. (2022). Google Trends data about mental health during COVID-19 pandemic using time series regression. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 125–129). IEEE. https://doi.org/10.1109/ISRITI56927.2022.10052965

Warnes, G. R., Bolker, B., Lumley, T., Magnusson, A., Venables, B., Ryodan, G., & Moeller, S. (2023, November 20). *gtools: Various R programming tools*. Comprehensive R Archive Network (CRAN). https://cran.r-project.org/package=gtools

Wasserman, D., Sokolowski, M., & Carli, V. (2020). Suicide. In E. Ryznar, A. B. Pederson, M. A. Reinecke, & J. G.

Csernansky (Eds.), *Landmark papers in psychiatry* (pp. 363–380). Oxford University Press. https://doi.org/10.1093/med/9780198836506.003.0022

Zhao, L., Lee, S., & Jeong, S. P. (2021). Decision tree application to classification problems with boosting algorithm. *Electronics*, *10*(16), 1–13. https://doi.org/10.3390/electronics10161903

APPENDICES

Table 1 Definition of Research Variables

Variable	Values
Suicide Attempt	1 = Suicide
(Y)	2 = Failed Suicide
Gender	0 = Unknown
(X_1)	1 = Man
	2 = Woman
Age	0 = Unknown
(X_2)	1 = Children (Less than 17 years old)
	2 = Adolescent (17-19 years old)
	3 = Young Adult (20-39 years old)
	4 = Mature (40-59 years old)
	5 = Elderly (60 years and above)
Rescue Action	1 = Available
(X_3)	2 = Not Available
Attempt Method	1 = Hanging Oneself
(X_4)	2 = Taking Drugs, Poisons, or Dangerous Chemicals
	3 = Cutting the Neck or Abdomen with a Knife
	4 = Burning Oneself
	5 = Jumping from a Bridge, Building, or Tall Building
	6 = Jumping into a River, Lake, Pond, Sea, or Well
	7 = Crashing into a Train
	8 = Other Methods
Mental Disorder	0 = Unknown
(X_5)	1 = Depression (Tendency to be Hopeless or Frustrated)
	2 = Bipolar Disorder (Tendency to be Quiet)
	3 = Schizophrenia (Tendency to Hallucinate)
	4 = Anxiety Disorder (Anxious or Panicky)
Trigger	0 = Unknown
Risk	1 = Debt, Online Loans, or Financial Problems
(X_6)	2 = Breakup, Victim of Cheating, or Romance Problems
	3 = Divorce
	4 = Killing a Spouse or Family
	5 = Being a Criminal Suspect
	6 = Prolonged Illness
	7 = Following a Death in the Family
	8 = Job Loss (Layoff)
	9 = Problems at Work or Business is Slow
	10 = Gambling or Online Gambling
	11 = Conflict with Family, Spouse, or Children
	12 = Other Triggers

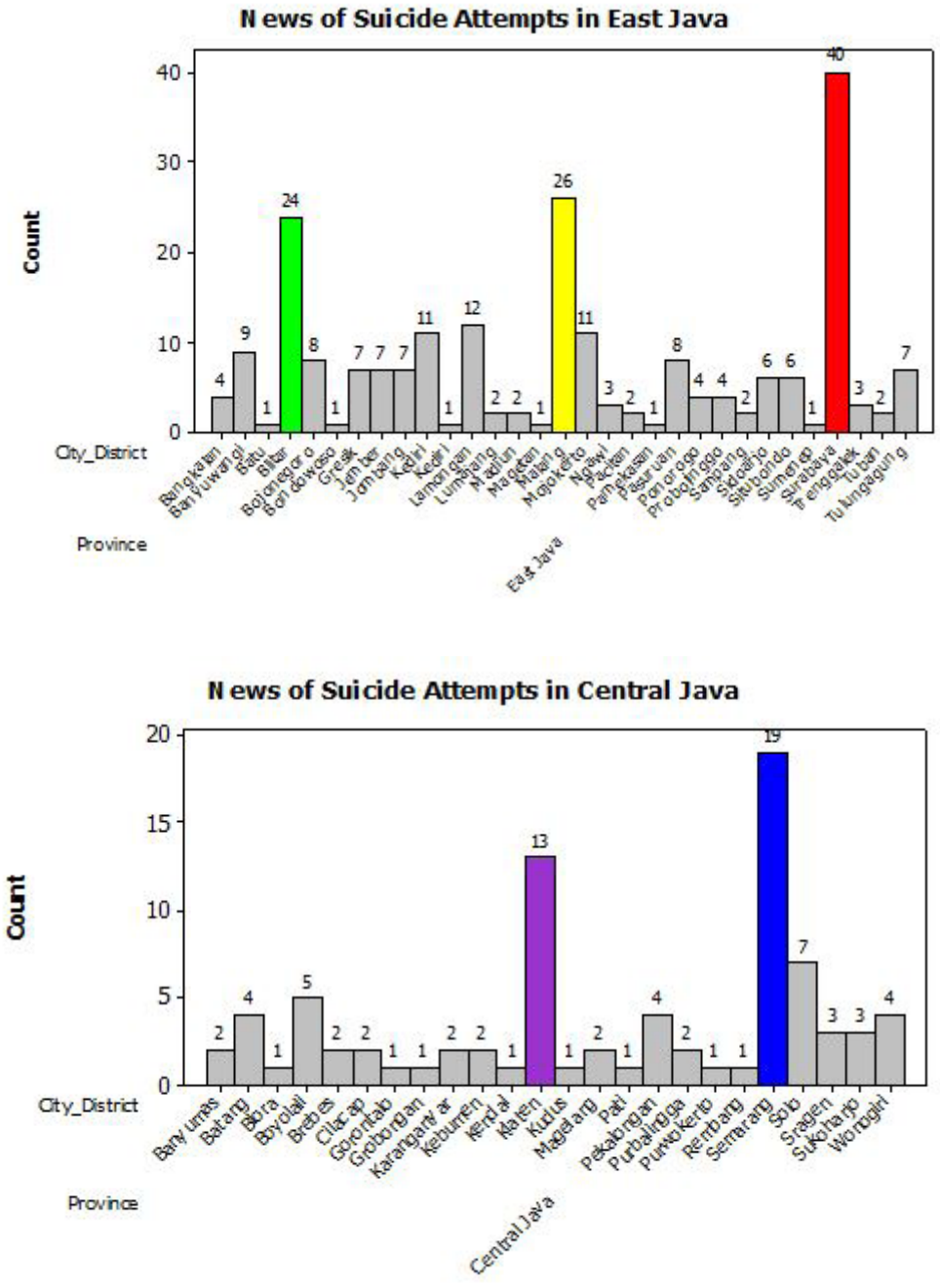

Figure 1 The Number of Suicide Attempt News in East Java
and Central Java Observed from January 2022 to July 2024

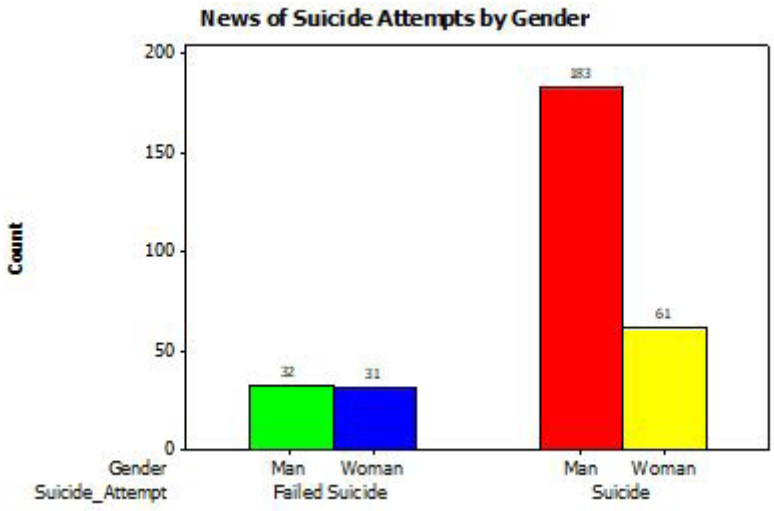

Figure 2 Number of Suicide Attempt News by Gender.

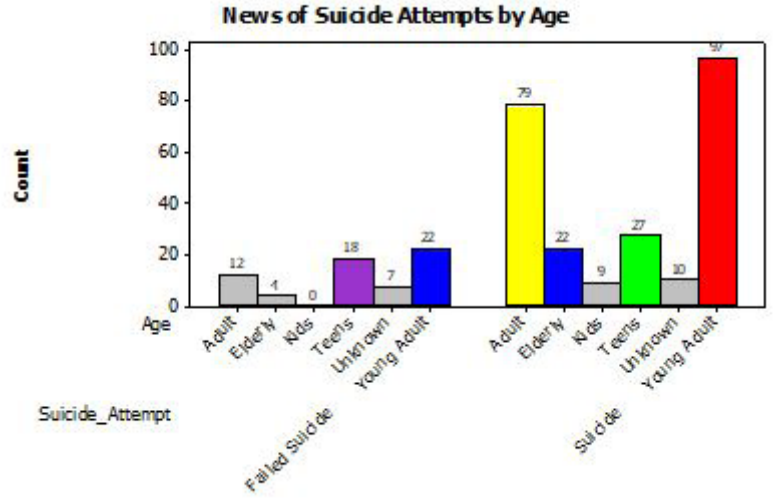

Figure 3 Number of Suicide Attempt News by Age

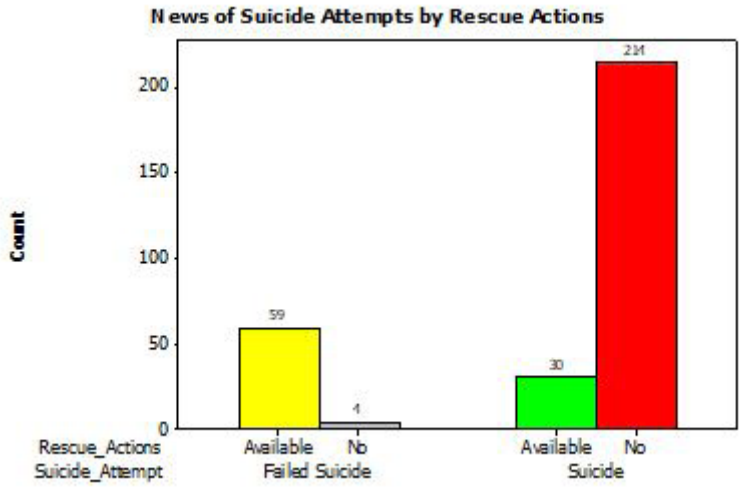

Figure 4 Number of Suicide Attempt News Based on Rescue Action

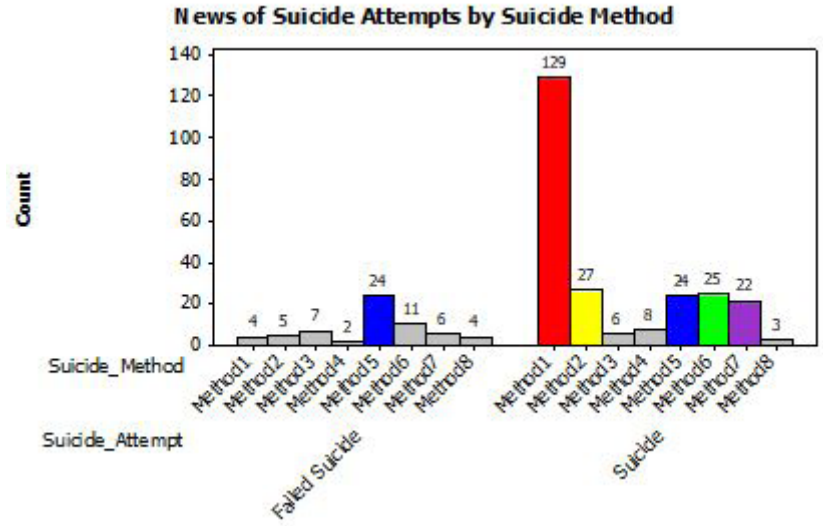

Figure 5 Number of Suicide Attempt News Based on the Attempt Method

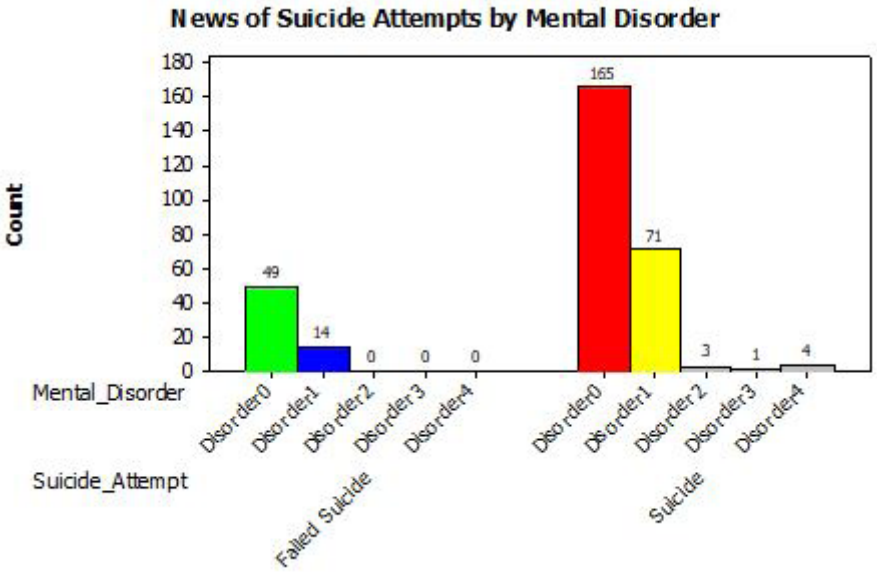

Figure 6 Number of Suicide Attempt News Based on Mental Disorder

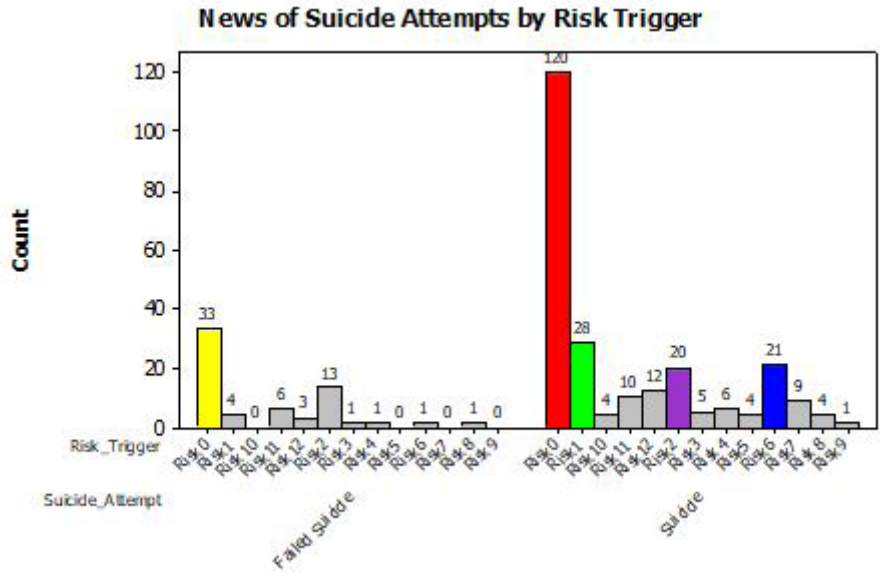

Figure 7 Number of Suicide Attempt News by Trigger Risk

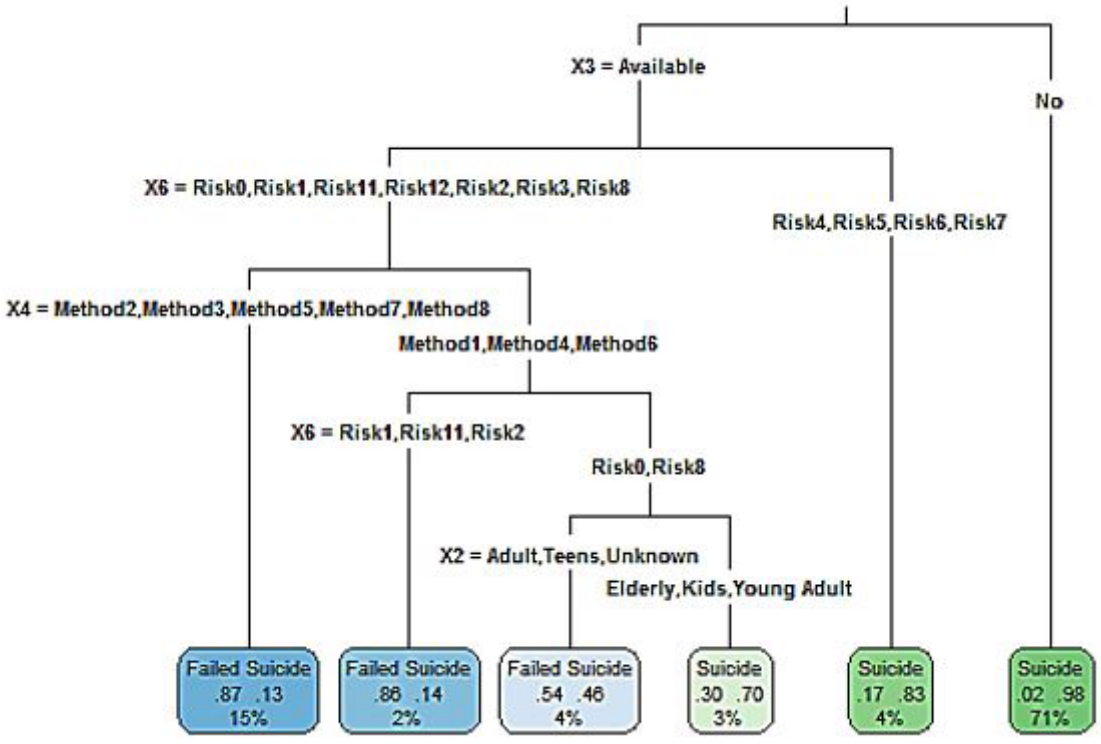

Figure 8 Decision Rules in the First Subset ($X_2 X_3 X_4 X_6$)

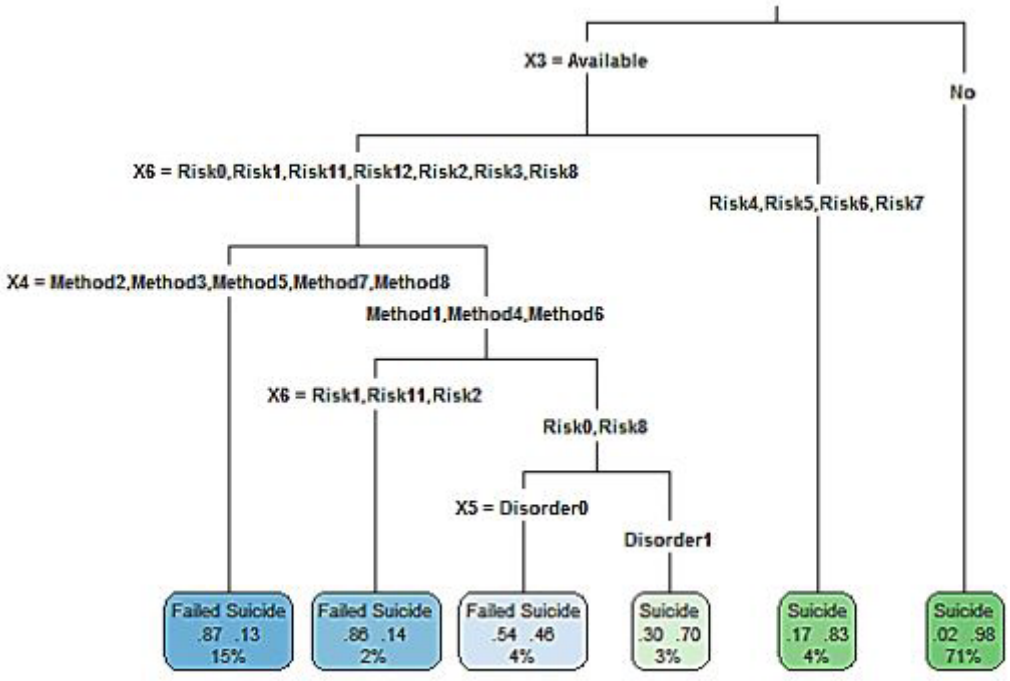

Figure 9 Decision Rules in the Second Subset ($X_3X_4X_5X_6$)

Table 2 All Possible Subsets Formed to Find the Best Subset

Possible Subset	Accuracy	Kappa	F1-Score
$Y = X_1X_2$	0.8046	0.2509	0.3478
$Y = X_1X_3$	0.8893	0.7056	0.7763
$Y = X_1X_4$	0.8208	0.3184	0.4086
$Y = X_1X_5$	0.7948	0.0000	NA
$Y = X_1X_6$	0.8046	0.1783	0.2500
$Y = X_2X_3$	0.8893	0.7056	0.7763
$Y = X_2X_4$	0.8208	0.2871	0.3678
$Y = X_2X_5$	0.7948	0.0000	NA
$Y = X_2X_6$	0.8013	0.0992	0.1408
$Y = X_3X_4$	0.9121	0.7288	0.7840
$Y = X_3X_5$	0.8925	0.6645	0.7317
$Y = X_3X_6$	0.9153	0.7601	0.8143
$Y = X_4X_5$	0.8078	0.3620	0.4779
$Y = X_4X_6$	0.8306	0.3694	0.4583
$Y = X_5X_6$	0.7948	0.0000	NA
$Y = X_1X_2X_3$	0.8990	0.7027	0.7669
$Y = X_1X_2X_4$	0.8306	0.4035	0.5000
$Y = X_1X_2X_5$	0.8046	0.2509	0.3478
$Y = X_1X_2X_6$	0.8143	0.2723	0.3596
$Y = X_1X_3X_4$	0.9121	0.7288	0.7840
$Y = X_1X_3X_5$	0.8925	0.6645	0.7317
$Y = X_1X_3X_6$	0.9153	0.7601	0.8143
$Y = X_1X_4X_5$	0.8274	0.3024	0.3765
$Y = X_1X_4X_6$	0.8469	0.4982	0.5913

Possible Subset	Accuracy	Kappa	F1-Score
$Y = X_1X_5X_6$	0.8176	0.1941	0.2432
$Y = X_2X_3X_4$	0.9153	0.7548	0.8088
$Y = X_2X_3X_5$	0.9023	0.7232	0.7857
$Y = X_2X_3X_6$	0.9153	0.7601	0.8143
$Y = X_2X_4X_5$	0.8339	0.3389	0.4138
$Y = X_2X_4X_6$	0.8469	0.5279	0.6240
$Y = X_2X_5X_6$	0.7948	0.0000	NA
$Y = X_3X_4X_5$	0.9186	0.7706	0.8227
$Y = X_3X_4X_6$	0.9283	0.7829	0.8281
$Y = X_3X_5X_6$	0.9153	0.7601	0.8143
$Y = X_4X_5X_6$	0.8306	0.3694	0.4583
$Y = X_1X_2X_3X_4$	0.9153	0.7548	0.8088
$Y = X_1X_2X_3X_5$	0.9023	0.7232	0.7857
$Y = X_1X_2X_3X_6$	0.9153	0.7601	0.8143
$Y = X_1X_2X_4X_5$	0.8371	0.3936	0.4792
$Y = X_1X_2X_4X_6$	0.8534	0.5426	0.6341
$Y = X_1X_2X_5X_6$	0.8143	0.2723	0.3596
$Y = X_1X_3X_4X_5$	0.9186	0.7706	0.8227
$Y = X_1X_3X_4X_6$	0.9283	0.7829	0.8281
$Y = X_1X_3X_5X_6$	0.9218	0.7631	0.8125
$Y = X_1X_4X_5X_6$	0.8469	0.4982	0.5913
$Y = X_2X_3X_4X_5$	0.9186	0.7706	0.8227
$Y = X_2X_3X_4X_6$	0.9283	0.7854	0.8308
$Y = X_2X_3X_5X_6$	0.9153	0.7601	0.8143
$Y = X_2X_4X_5X_6$	0.8534	0.5255	0.6154
$Y = X_3X_4X_5X_6$	0.9283	0.7854	0.8308
$Y = X_1X_2X_3X_4X_5$	0.9186	0.7706	0.8227
$Y = X_1X_2X_3X_4X_6$	0.9283	0.7854	0.8308
$Y = X_1X_2X_3X_5X_6$	0.9218	0.7631	0.8125
$Y = X_1X_2X_4X_5X_6$	0.8599	0.5409	0.6261
$Y = X_1X_3X_4X_5X_6$	0.9283	0.7854	0.8308
$Y = X_2X_3X_4X_5X_6$	0.9283	0.7854	0.8308
$Y = X_1X_2X_3X_4X_5X_6$	0.9283	0.7854	0.8308

Table 3 Possible Subset Selected as the Best Subset

Possible Subset	Accuracy	Kappa	F1-Score	Tree that Formed
$Y = X_3X_6$	0.9153	0.7601	0.8143	X_3X_6
$Y = X_3X_4X_6$	0.9283	0.7829	0.8281	$X_3X_4X_6$
$Y = X_2X_3X_4X_6$	0.9283	0.7854	0.8308	$X_2X_3X_4X_6$
$Y = X_3X_4X_5X_6$	0.9283	0.7854	0.8308	$X_3X_4X_5X_6$
$Y = X_1X_2X_3X_4X_6$	0.9283	0.7854	0.8308	$X_2X_3X_4X_6$
$Y = X_1X_3X_4X_5X_6$	0.9283	0.7854	0.8308	$X_3X_4X_5X_6$
$Y = X_2X_3X_4X_5X_6$	0.9283	0.7854	0.8308	$X_2X_3X_4X_6$
$Y = X_1X_2X_3X_4X_5X_6$	0.9283	0.7854	0.8308	$X_2X_3X_4X_6$