Bulletin of Social Informatics Theory and Application
Vol. 5, No. 1, March 2021, pp. 38-46
ISSN 2614-0047

Bayes interpretation for smoke-free area cities index

Tazkiyah Herdi [1,*], Ardiansyah Dores [2]

[a] Universitas Mercu Buana, Jakarta, Indonesia
[1] tazkiyah.herdi@mercubuana.ac.id*; [2] ardian@mercubuana.ac.id
* corresponding author

ARTICLE INFO

Article history
Received December 17, 2020
Revised January 16, 2021
Accepted January 30, 2021

Keywords
Smoke-free policy
Decision support system
Naïve bayes classifier

ABSTRACT

To control tobacco consumption in Indonesia, whose prevalence increased from 32.8% to 33.8%, local governments issued regional regulations on Smoke-Free Area Policy with support from the central government. Until 2019, there were still 166 cities/districts governments that were yet to issue the regulation out of 514 cities/districts. The increase in the number of active smokers and individuals still exposed to cigarette smoke shows that efforts have not been optimized to reduce tobacco consumption. Furthermore, no control effort has been discovered regarding the level of success of the policies that have been applied. Therefore, this research discusses the surveys carried out in cities/Districts that have applied the smoke-free policy, using indicators such as ideal questions relating to the policy. Naïve Bayes Classifier is one of the Decision Support System (DSS) classification methods used to classify the survey results into good, fair or poor categories to determine whether each city/district has implemented the issued regulation. Based on the results from the classification of the three cities/districts using the classifier, Bogor Regency was classified as good while Lombok and Padang Cities were classified as poor.

1. Introduction

According to the Global Youth Tobacco Survey carried out in 2019, 19.2% of schoolchildren were smoking, comprised 35.6% male and 3.5% female. Meanwhile, 57.8% of schoolchildren aged 13-15 years were exposed to cigarette smoke at home and 66.2% in public places, implying that 6 out of every ten school children aged 13-15 years were exposed to cigarette smoke at home in public places [1]. The prevalence of smoking among adolescents between the ages of 10-18 years increased continuously from 7.2% in 2013 to 9.1% in 2018. Furthermore, tobacco consumption among the population aged above 15 years also increased from 32.8% to 33.8%.

There is growing international interest in advancing 'the tobacco endgame. A previous study used New Zealand smoke-free goal for 2025 as an example to model the impacts on smoking prevalence, health gains, and cost savings. Furthermore, the cost savings consist of strategies, which are 10% annual tobacco tax increases, a tobacco-free generation, a substantial outlet reduction strategy, and a sinking lid on tobacco supply [2]. In terms of controlling tobacco consumption, the government should (1) increase the cigarette tax [3], (2) make policy for TAPS (tobacco, advertisement, promotion, and sponsor) ban [4], (3) Implement a warning on the effects of smoking with PHW (Pictorial Health Warning) on cigarette advertisements, and packaging, (4) make policy on smoke-free area, which are places or areas that are declared prohibited from smoking, producing, introducing and/promoting tobacco products.

The World Health Organization (WHO) in the Framework Convention on Tobacco Control (FCTC) stated that both the local and central government should enforce and implement the regulation in the jurisdiction of their respective countries as stipulated in national law and actively promote the enactment and application of legislative, executive, administrative and action measures. In order to

https:doi.org/10.31763/businta.v5i1.286

protect against exposure to cigarette smoke in workplaces, indoors, public transport, closed public places, and in other public places [5], WHO introduced the MPOWER concept (Monitor, Protect, Offer, Warn, Enforce, Raise). It is an effective measure to assist in the implementation at the country level to reduce tobacco demand. The P stands for Protect, which is education on protection. It also emphasizes that clean air is a human right. Public support is the critical factor for the success of smoke-free laws, which is obtainable through effective education about the dangers of passive smoking and a clear explanation of policy objectives.

To control active smoking in Indonesia, the city and district government issued local regulations on smoke-free policy and TAPS ban with support from the central government. Until 2019, there were still 166 cities/districts that were yet to issue the regulation. The increasing number of active smokers and individuals exposed to cigarette smoke shows that efforts have not been optimized to reduce tobacco consumption. Furthermore, no control effort has been discovered regarding the success of policies applied.

Previous studies have been made for the social impacts of this smoke-free policy. In Jayapura, researchers monitored the compliance and barriers to applying the smoke-free area and social studies law in Jayapura. They found that the smoke-free area criteria compliance following government regulations was only 17% [6]. Smoke-free policy implementation could be enhanced with information about second-hand smoke exposure (SHSe) for smokers and non-smokers [7]. In Palembang and Bogor, many respondents find smoke-free policy necessary. However, they still experienced of asking smokers not to smoke in restricted areas [8]. The role of campus smoking policies on reducing student smoking behavior was concerned across US. Other types of policies concerned included partial smoking restriction and integration of preventive education and/or smoking cessation programs into college-level policies. The results of the studies had reviewed that policies were found to significantly reduce smoking behavior and pro-smoking attitudes over time [9]. Measurement of the level of implementation of smoke-free area policy and TAPS (Tobacco Advertisements Promotions and Sponsors) Ban has been calculated in previous research. The results showed that smoke-free area policy and TAPS Ban using the SAW (Simple Additive Weighting) analysis to weigh all applicable indicators is capable of showing the ranking value [10].

According to Arnot and Pervan, Decision Support System (DSS) has been used in many types of research, which focused on the evolution of several sub-groupings of research and practice such as personal, group, negotiation, intelligence, and knowledge management based decision support system. The others include executive information systems/business intelligence and data warehousing" [11]. Decision support systems aids human cognitive deficiencies by integrating various sources of information, providing intelligence, access to relevant knowledge, and aiding decision-making [12], [13]. A decision is a choice among several alternatives, and Decision Making refers to the whole process of assessing the problem, collecting and verifying information, identifying alternatives, anticipating consequences of decisions, making a choice using sound and logical judgment based on available information, informing others of the decision and rationale, evaluating decisions. A multi-attribute decision is applied for an assessment mechanism to get recommendations from prospective corporate employees with high benefits and low costs for the company [14], [15].

One of the DSS classification methods is the Naive Bayes Classifier, which was derived from Bayes' theorem. It is a statistical calculation that is carried out by evaluating the probability of similarity of an existing case on a case-by-case basis. It has a high degree of accuracy and speed when applied to large databases [16], [17]. In addition, it allows researchers to formally incorporate prior information, whether qualitative or quantitative, into their analyses.

Bayes' Theorem developed over time into Naïve Bayes classification method. Naïve Bayes classifier is applied to the text classification for news categories, which explains the difficulty of newsreaders choosing online news according to their categories. This method can be applied to classify online news portals in finance, lifestyle, news, and sports [18]. Furthermore, it is applicable in the government and agribusiness sectors. Its Algorithm is also applicable in regional government performance assessment on Regional Budgeting Management since there is no control system in the regional government to manage the Budget properly. Therefore, an assessment system using this classification may be carried out to determine the performance of the village government in managing the budgets. Meanwhile, in the agribusiness sector, research also proposed a shallots classification method based on quality using the Naïve Bayes Classifier, which classified them into three classes: good, medium, and poor quality [19].

Naive Bayes Classifier has been developed, and it is a method of calculating the decision support system assessment in various industrial sectors. This research discusses the surveys carried out in Cities/Districts that have implemented smoke-free policies. This method was used to classify the survey results from each city/district to determine whether the implemented smoke-free policy is in good, fair, or poor categories. This research focuses on data samples in Lombok and Padang Cities, Bogor District to get the appropriate Bayes method calculation formula.

2. Method

Fig. 1 presents the research framework. Literature reviews in books, journals, regional regulations, research reports, and relevant information were carried out to determine the method to be used. Furthermore, data collection was carried out through the following activities: collaboration between the author's team and non-governmental organization to formulate assessment indicators by global regulations, which are the Framework Convention on Tobacco Control (FCTC), government and local regulations and indicators for assessing child-friendly cities. After the indicators were determined, several questions were asked concerning the indicators to be answered by the communities. In addition, the indicator survey was tested in the areas of Padang city, Lombok city, and Bogor district to obtain sample data.

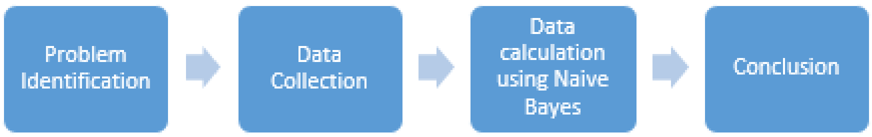

Fig. 1. The research framework

Table.1 Places of smoke-free areas based on regulations

	Smoke-free area policy		
No	**FCTC**	**Child-friendly city**	**Government Regulation no. 109 year 2012**
1	Indoor office space	Health service	Health service
2	Public transportation	Education places	Education places
3	Indoor public places	Public places	Child-friendly indoor and outdoor
4	Other public places	Child-friendly indoor and outdoor	Worship places
5		Worship places	Public transportation
6		Public transportation	Office spaces
7			Public and other determined places.

After the survey data was collected, a classification calculation simulation using the Bayes Theorem method was carried out to ascertain whether the smoke-free policy implemented by the city/district is reasonable, fair, or poor. The calculation, conclusion, and results were generated after going through these stages.

Fig. 2 shows the Naïve Bayes Learning process for training data. The first step was preprocessing, the survey documents were used to process the cleaning and grouping according to cities in good, adequate or insufficient categories. For the second step, each type of data on the training data variables was searched. When it is available, it was added to the frequency of that category. However, the data is added as a new one when the searching failed

The probability of (P (wk | vj)) was calculated, where wk is the indicator variable of smoke free areas and vj are the categories: good, fair, and poor. Furthermore, the amount of survey data that has a conformity value was added and the value of (P (vj)) was calculated, Therefore, the probability of (P (vj) was recalculated

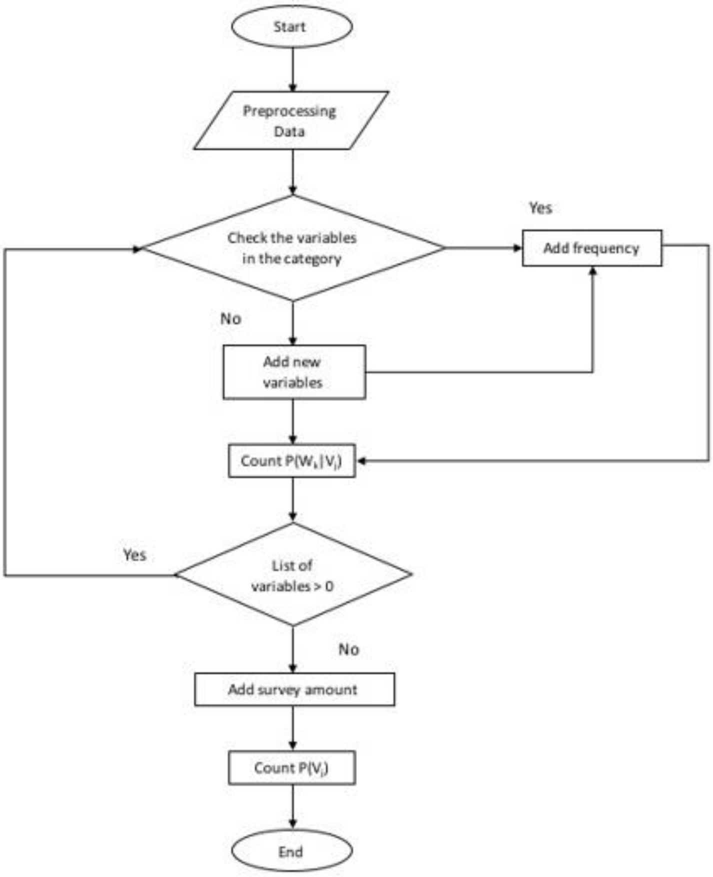

Fig. 2. Bayes Learn Process Flowchart

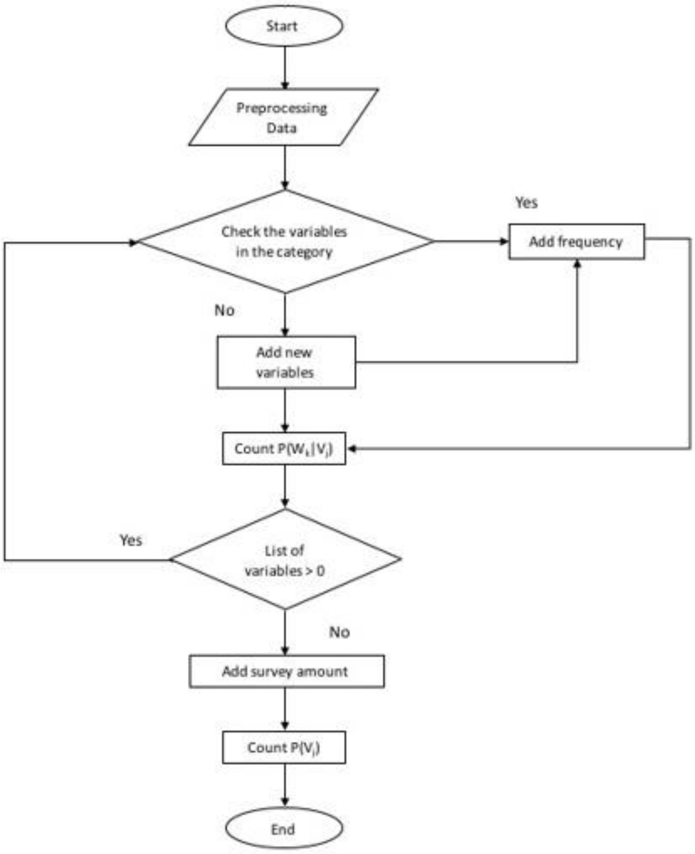

Fig. 3. Bayes Classifying Process Flowchart

The purpose of the Naïve Bayes classifying process is to discover the highest possible value to make a test data classification value in the correct category. Fig. 2 shows the design for the classification of the test data, which was carried out by checking the variables in each category with the highest frequency. Furthermore, the probability value of (P (wk | vj)) was calculated as well as the probability value of (P (vj)) * (P (wk | vj)). The next stage was the result in the second step process, each category was compared for the most outstanding value, and the document were included in that category.

3. Results and Discussion

The number of survey data for each city was grouped into good, fair, and poor categories. The good category represents the answers of respondents that high-frequently see no-smoking signs and officers warning the smokers in particular places while low-frequently see people smoking, cigarette butts, and TAPS in particular places. Meanwhile, the Fair category represents the answers of respondents that med-frequently see no-smoking signs and officers that warned smokers at particular places while med-frequently saw smoking people, cigarette butts, and TAPS in particular places. Also, the poor category represents the answers of respondents that low-frequently see no-smoking signs and officers that warned the smoker in that place while, high-frequently see smoking people, cigarette butts, and TAPS in particular places.

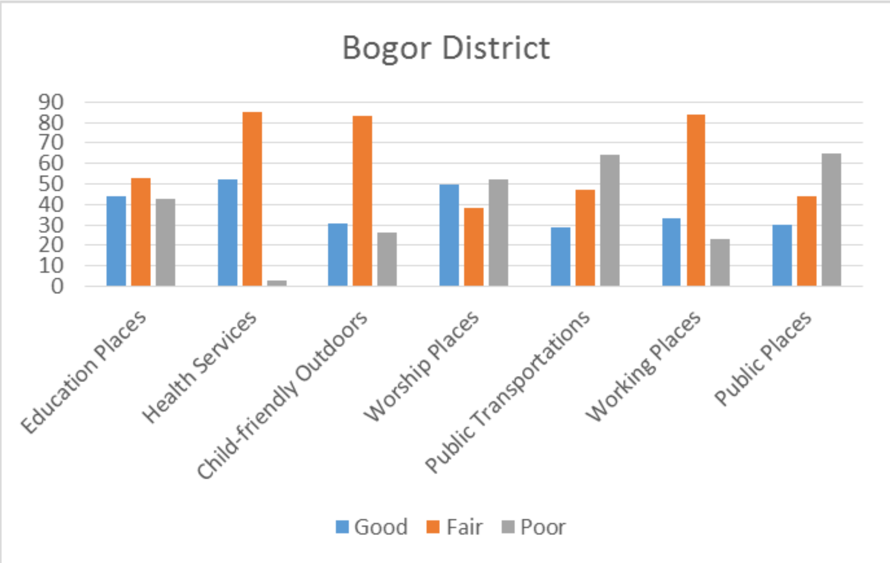

Fig. 4. Number of frequencies based on respondents' answers in Bogor District

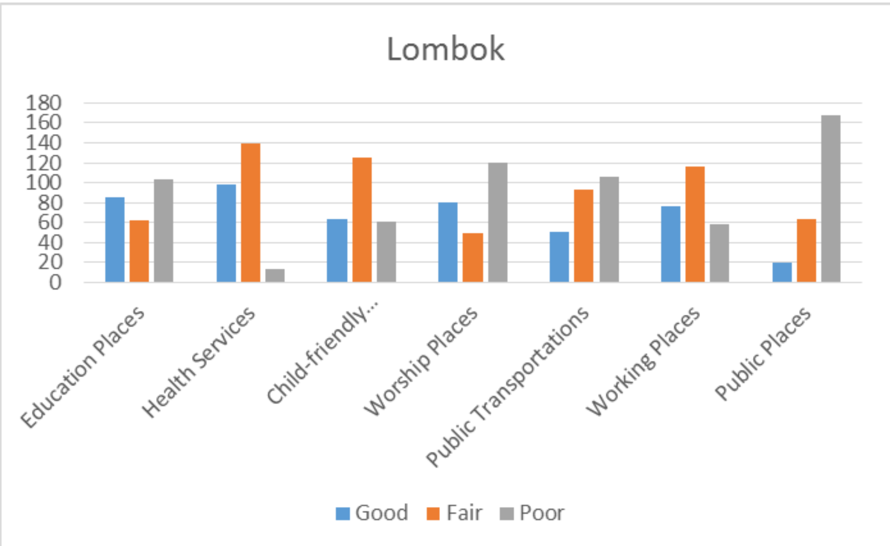

Fig. 5. Number of frequencies based on respondents' answers in Lombok

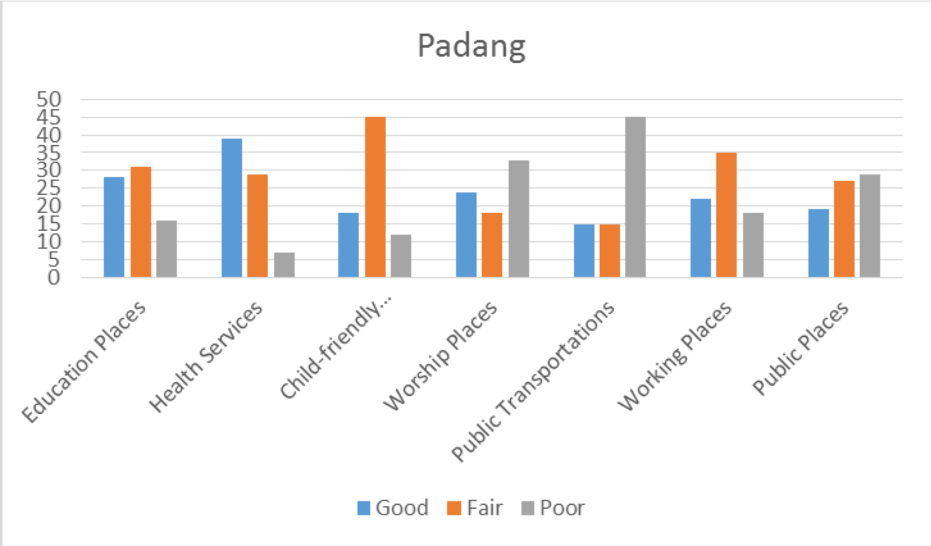

Fig. 6. Number of frequencies based on respondents' answers in Padang

In summary, the graphs show that the frequency of good and fair categories mostly took place in health services, followed by education places and child-friendly outdoors. Meanwhile, the poor category took place mainly in public places and public transportations

The health services in Bogor district (Fig. 4) took seriously about smoke-free area policy, because it had the highest frequency of good category, at roughly over 50. There was a slight difference between health services and child-friendly outdoors. The highest frequency of appropriate category was located in health services, roughly above 80. Meanwhile, the highest frequency of poor category was located in public transportations and public places.

Furthermore, In Lombok (Fig. 5), the highest frequency of excellent and fair categories was also located in health services, at about 140 and nearly 100. Meanwhile, the public places had the highest frequency of poor category. In Padang (Fig. 6), health services, child-friendly outdoors, and public transportations had the highest good, fair and poor frequency, respectively.

Table 2 shows the Bayes training data which used all respondents' answers from three cities. Education places, health services, child-friendly outdoors, places of worship, public transportations, workplaces, and public places were assumed as variables. Then the categories were good, fair and poor.

Table.2 Frequency of Variables to Categories

No	Variables	Good	Fair	Poor	Total
1	Education Places	157	146	162	465
2	Health Services	189	253	23	465
3	Child-friendly Outdoors	113	253	99	465
4	Worship Places	154	106	205	465
5	Public Transportations	95	155	215	465
6	Working Places	131	235	99	465
7	Public Places	70	134	261	465
	Total Data Training	909	1282	1064	3255

After being grouped based on the place and frequency of categories, the probability between variables to the category was searched with the formula of the equation $P(W_k|V_j) = \frac{n_k+1}{n+|vocabulary|}$ nk was the category frequency in a variable, and n is all frequencies in all variables in a category. Vocabulary are the frequency of all variables from all categories. Table 3 shows the calculation of the probability of each variable to categories. This formula is used as the training data for data testing in Table 4, 5, and 6. Table 4 shows that the Bogor Regency testing document is in the Fair category. Based on the probability values shown in Table 5, the Lombok City testing document is in the Poor category. Finally, Table 6 shows that the Padang City testing document is in the Poor category.

Table.3 Probability Variables to Categories as Training Data

No	Variables	P[Variables|Good]	P[Variables|Fair]	P[Variables|Poor]
1	Education Places	0.037944284	0.032400264	0.037740218
2	Health Services	0.045629203	0.05598413	0.005556842
3	Child-friendly Outdoors	0.027377522	0.05598413	0.023153508
4	Worship Places	0.037223823	0.023583866	0.047696226
5	Public Transportations	0.023054755	0.034383954	0.050011577
6	Working Places	0.031700288	0.052016751	0.023153508
7	Public Places	0.017050913	0.029755345	0.06066219

Table.4 Data Testing for Bogor District

	Bogor District	**Frequency**			**P[vj] * P[wk |vj]**		
No	**Variables**	**Good**	**Fair**	**Poor**	**Good**	**Fair**	**Poor**
1	Education Places	44	53	43	*	0.010800088	*
2	Health Services	52	85	3	*	0.018661377	*
3	Child-firendly outdoors	31	83	26	*	0.018661377	*
4	Worship Places	52	38	50	0.012407941	*	*
5	Public Transportations	29	47	64	*	*	0.016670526
6	Working Places	33	84	23	*	0.017338917	*
7	Public Places	31	44	65	*	*	0.02022073
	Total				0.012407941	0.065461759	0.036891256

Table.5 Data Testing for Lombok City

	Lombok City	Frequency			P[vj] * P[wk |vj]		
No	Variables	Good	Fair	Poor	Good	Fair	Poor
1	Education Places	85	62	103	*		0.012580073
2	Health Services	98	139	13	*	0.018661377	
3	Child-friendly outdoors	64	125	61	*	0.018661377	
4	Worship Places	80	50	120	*	*	0.015898742
5	Public Transportations	51	93	106	*	*	0.016670526
6	Working Places	76	116	58	*	0.017338917	
7	Public Places	20	63	167	*	*	0.02022073
	Total				*	0.054661671	0.06537007

Table.6 Data Testing for Padang City

No	Padang City	Frequency			P[vj] * P[wk |vj]		
	Variables	Good	Fair	Poor	Good	Fair	Poor
1	Education Places	28	31	16	*	0.010800088	*
2	Health Services	39	29	7	0.015209734		*
3	Child-friendly outdoors	18	45	12	*	0.018661377	*
4	Worship Places	24	18	33	0.012407941	*	0.015898742
5	Public Transportations	15	15	45	*	*	0.016670526
6	Working Places	22	35	18	*	0.017338917	
7	Public Places	19	27	29	*	*	0.02022073
	Total				0.027617675	0.046800382	0.052789998

4. Conclusion

To conclude, the frequency of excellent and fair categories mainly was located in health services, followed by education places and child-friendly outdoors. Meanwhile, the poor category was primarily located in public places and public transportations. Documents for data testing include 3 cities that were assessed separately to be classified into a category. The classification results using the Naïve Bayes classifier include Bogor Regency as good, Lombok and Padang cities as poor. Furthermore, this classification helps analyze and design a decision support system to assess all cities/districts implementing the local policy of smoke-free areas. Other decision-making calculation methods should be tried out to obtain results following the policy stakeholders.

Acknowledgment

The authors express gratitude to Universitas Mercu Buana for funding this research through research scheme project 02-5/209/B-SPK/III/2020. Furthermore, acknowledge the use of service and facilities of the Business Intelligence Laboratory at Department of Information System, Universitas Mercu Buana.

References

[1] World Health Organization, “2019 GYTS Fact Sheet Indonesia,” 2019.

[2] F. S. van der Deen *et al.*, “Impact of five tobacco endgame strategies on future smoking prevalence, population health and health system costs: two modelling studies to inform the tobacco endgame,” *Tob. Control*, vol. 27, no. 3, pp. 278–286, May 2018.

[3] L.-M. Ho, C. Schafferer, J.-M. Lee, C.-Y. Yeh, and C.-J. Hsieh, “Raising cigarette excise tax to reduce consumption in low-and middle-income countries of the Asia-Pacific region:a simulation of the anticipated health and taxation revenues impacts,” *BMC Public Health*, vol. 18, no. 1, p. 1187, Dec. 2018.

[4] D.-J. A. van Mourik, G. E. Nagelhout, M. C. Willemsen, B. van den Putte, and H. de Vries, “Differences in smokers’ awareness of the health risks of smoking before and after introducing pictorial tobacco health warnings: findings from the 2012–2017 international tobacco control (ITC) Netherlands surveys,” *BMC Public Health*, vol. 20, no. 1, p. 512, Dec. 2020.

[5] WHO FCTC, “WHO Framework Convention on Tobacco Control,” 2003.

[6] W. Wahyuti, S. Hasairin, S. Mamoribo, A. Ahsan, and D. Kusuma, “Monitoring Compliance and Examining Challenges of a Smoke-free Policy in Jayapura, Indonesia,” *J. Prev. Med. Public Heal.*, vol. 52, no. 6, pp. 427–432, Nov. 2019.

[7] J. Anthony *et al.*, “Qualitative Assessment of Smoke-Free Policy Implementation in Low-Income Housing: Enhancing Resident Compliance,” *Am. J. Heal. Promot.*, vol. 33, no. 1, pp. 107–117, Jan. 2019.

[8] M. R. Kaufman, A. P. Merritt, R. Rimbatmaja, and J. E. Cohen, "'Excuse me, sir. Please don't smoke here'. A qualitative study of social enforcement of smoke-free policies in Indonesia," *Health Policy Plan.*, vol. 30, no. 8, pp. 995–1002, Oct. 2015.

[9] B. L. Bennett, M. Deiner, and P. Pokhrel, "College anti-smoking policies and student smoking behavior: a review of the literature," *Tob. Induc. Dis.*, vol. 15, no. 1, p. 11, Dec. 2017.

[10] T. Herdi, A. Dores, and F. Masya, "Simple Additive Weighting in Selection of Cities Performance of Non Smoking Area and Tobacco Advertisements, Promotions and Sponsors Ban," *Int. J. Inf. Syst. Comput. Sci.*, vol. 4, no. 2, pp. 71–79, 2020.

[11] D. Arnott and G. Pervan, "A critical analysis of decision support systems research," in *Formulating Research Methods for Information Systems*, London: Palgrave Macmillan UK, 2015, pp. 127–168.

[12] S. F. Crone, S. Lessmann, and R. Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing," *Eur. J. Oper. Res.*, vol. 173, no. 3, pp. 781–800, 2006.

[13] G. E. Phillips-Wren, O. M. Ferreiro, G. Forgionne, and H. Desai, "Adoption of Decision Support Systems (DSS) in a Developing Country," *J. Decis. Syst.*, vol. 16, no. 4, pp. 425–449, 2007.

[14] R. Accorsi, R. Manzini, and F. Maranesi, "A decision-support system for the design and management of warehousing systems," *Comput. Ind.*, vol. 65, no. 1, pp. 175–186, Jan. 2014.

[15] B. A. Alyoubi, "Decision Support System and Knowledge-based Strategic Management," *Procedia Comput. Sci.*, vol. 65, pp. 278–284, 2015.

[16] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.

[17] K. I. Qazi, H. K. Lam, B. Xiao, G. Ouyang, and X. Yin, "Classification of epilepsy using computational intelligence techniques," *CAAI Trans. Intell. Technol.*, vol. 1, no. 2, pp. 137–149, 2016.

[18] A. Dores, F. Masya, and H. Prastiawan, "Indonesian Text News Classification Using the Naive Bayes Algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 7, no. 8, pp. 159–169, 2018.

[19] A. Susanto, Z. H. Dewantoro, C. A. Sari, D. R. I. M. Setiadi, E. H. Rachmawanto, and I. U. W. Mulyono, "Shallot Quality Classification using HSV Color Models and Size Identification based on Naive Bayes Classifier," *J. Phys. Conf. Ser.*, vol. 1577, p. 012020, Jul. 2020.