



Computer Methods and Programs in Biomedicine Update

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update



Why do youths initiate to smoke? A data mining analysis on tobacco advertising, peer, and family factors for Indonesian youths

Enny Rachmani^{a,*}, Sri Handayani^a, Kriswiharti Kun Saptorini^a, Nurjanah^a, Dian Kusuma^b, Abdillah Ahsan^c, Edi Jaya Kusuma^a, Suleman Atique^d, Jumanto Jumanto^e

^a Faculty of Health Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

^b Department of Health Services Research and Management, School of Health & Medical Sciences, City St George's University of London, London, UK

^c Faculty of Economics and Business, University of Indonesia, Depok 16424, Indonesia

^d Department of Public Health Science, Faculty of Landscape and Society, Norwegian University of Life Sciences (NMBU), Ås, Norway

^e Faculty of Humanities, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

ARTICLE INFO

Keywords:

Advertising
Family
Peer influence
Youth
Smoking
Indonesia
Data mining

ABSTRACT

Global Youth Tobacco Survey (GYTS), Indonesia showed that 60,9 % of students noticed cigarette advertisements or promotions in outdoor media. Our study aimed to understand the impact of outdoor tobacco advertising and peer and family association with Youth's smoking behavior.

This study deployed a cross-sectional approach to explore factors related to youth smoking behavior, such as peers, family, and tobacco advertising. The GYTS questionnaire was adapted as the instrument and distributed to 400 students from 20 high schools to observe smoking behavior. The chosen schools based on the previous study whose classify school in hot-spot and non hot-spot area. This study applied a data mining approach with a decision tree to generate the models.

This study generates a decision tree model that describes the peer factor as the key to introducing Youth to smoking. The model also reveals that youth in the non-hotspot advertising area are not likely to develop Youth to smoke. The model has a performance classification of 77.5 % This study found that youth with smoking fathers are more likely to start smoking earlier, youth whose both parents are smokers, and mothers who are smokers have a confidence level of 100 % to smoke. Further research is warranted to investigate rural districts to explore any regional and socioeconomic variations.

Introduction

The Global Youth Tobacco Survey (GYTS) reveals that >60 % of Indonesian students have been exposed to tobacco advertising in outdoor media, raising concerns. In response, our study investigated the impact of this advertising on youth's smoking behavior, examining its interplay with peer and family influences. This issue is worsened by Indonesia's health landscape, which faces a dual challenge: rising chronic diseases due to changing demographics and lifestyles, alongside alarmingly high rates of infectious diseases. The surge in Non-communicable Diseases (NCDs) is extraordinary, contributing significantly to global mortality, comprising nearly 70 % of deaths worldwide [1]. Tobacco consumption is accountable for 225,720 fatalities

annually, representing 14.7 % of all recorded deaths. Notably, the predominant cause of mortality attributed to tobacco is within the population having cardiovascular diseases (CVDs). Particularly concerning is the increased likelihood of tobacco-induced CVD among younger individuals [2].

Tobacco advertising stands out as one of the most assertive and innovative, employed worldwide for promoting tobacco products. It serves as a powerful tool in shaping perceptions, attitudes, and behaviors related to smoking [3]. Particularly in Indonesia, tobacco companies integrate advertising as a pivotal tool to cultivate new consumers, often targeting younger ones, thus embedding these promotional campaigns into the country's societal roots. [4]. Concurrently, Indonesia exhibits a deficiency in comprehensive national tobacco control

* Corresponding author at: Faculty of Health Science, Universitas Dian Nuswantoro, No 5-11, Nakula Street, 1st Floor, Building D, Semarang, Indonesia, 50131.
E-mail addresses: enny.rachmani@dsn.dinus.ac.id (E. Rachmani), sri.handayani@dsn.dinus.ac.id (S. Handayani), kriswiharti@dsn.dinus.ac.id (K.K. Saptorini), nurjanah@dsn.dinus.ac.id (Nurjanah), dian.kusuma@city.ac.uk (D. Kusuma), abdillah.hasan@ui.ac.id (A. Ahsan), edi.jaya.kusuma@dsn.dinus.ac.id (E.J. Kusuma), suleman.atique@nmbu.no (S. Atique), ilhamj@dsn.dinus.ac.id (J. Jumanto).

<https://doi.org/10.1016/j.cmpup.2024.100168>

measures. The existing Government Regulation (PP) No. 109/2012 about tobacco control, permits cigarette advertisements, promotions, and sponsorship across all media platforms [5]. As per the findings of the Global Youth Tobacco Survey (GYTS), Indonesia demonstrates that 19.2 % of students are engaged with various forms of tobacco products. Moreover, a significant majority, accounting for 60.9 % of students, reported exposure to cigarette advertisements or promotions in outdoor media [6]. This issue is rising because of the high density of tobacco advertising around the school areas [7,8]. The government should pay attention to this because the prevalence of child smokers in Indonesia between 2014 and 2019 increased marginally from 18.3 % to 19.2 % within the demographic aged 13 to 15 years old [9].

Tobacco advertisements in Indonesia display images related to friendships and helping to increase confidence and masculinity targeting the young age group [10]. It is related to young age's odds of smoking susceptibility [11]. Furthermore, peer smoking is one of the most significant factors increasing the possibility of smoking at a younger age [12,13]. Social learning theory introduced the term of social modeling, which refers to the perceived conduct of others as a primary source of influence on the observer's behavior such as smoking [14]. Studies have proven the relationship between peer smoking behavior and individual smoking, such that having friends who smoke increases the likelihood of becoming a smoker [15]. There is diverse evidence about peer perception of popularity and health behaviors, particularly smoking. For example, a study indicated that popular adolescents (aged 11–12) were more likely to smoke than their less popular peers. Other research, however, reveals that smoking may be connected with social isolation or having fewer friends (aged 11–19 years) and that popular individuals may be more influential than their less popular peers [16].

According to previous research, parental and peer factors are a significant component in adolescents' smoking behavior. Youth often imitate their parents' smoking habits, which can lead to heavy smoking in early adulthood [17]. The impact of parental modeling on youths' smoking behavior is widely recognized. Previous studies indicated that when at least one parent smokes, the risk of smoking uptake in childhood and adolescence increases almost twofold, and when both parents smoke, the risk nearly triples [18]. Adolescents are less likely to smoke if their parents are former or declining smokers. Quitting smoking earlier also reduces the likelihood of adolescent smoking.

The survey also showed that exposure to family smoking tends to persuade youth to smoke in the future and increase the odds of peer exposure. [13] Previous studies have shown the pattern of substance use, including tobacco use, related to perceived parenting style and peer consumption. [19] Student smokers are associated with deviant groups. [20]

This study followed the previous research in Semarang, Indonesia. [8,21] that aimed to examine the density and hotspots of outdoor tobacco advertisements near schools and investigate the relationship between youth smoking behavior and indicators of outdoor tobacco advertisement density and proximity in Indonesia, both of which were conducted in Semarang.

Our study attempted to explain the relationship between outdoor tobacco advertising, peer influence, family dynamics, and their interplay with youth's smoking behaviors. Employing a data mining methodology featuring a decision tree algorithm, we conducted a classification analysis to discern patterns among these factors. This research marks the primary utilization of a data mining approach to delve into the determinants of Youth's smoking behavior.

Materials and methods

Data collections

This study used a cross-sectional quantitative approach to investigate the determinants associated with youth smoking behavior, building upon a prior investigation focused on the density and proximity of

outdoor tobacco advertising in Indonesia. [8] The advertising data were acquired from our previous study conducted from November to December 2018 through a survey of outdoor tobacco advertisements. The advertisements included a wide range of media, such as billboards, videoboards, banners, retail signs, neon boxes, posters, and stickers. A total of 3453 outdoor tobacco advertisements, including those displayed in front of stores and retailers, were observed. The advertisements varied in size, ranging from small (roughly A4 size, measuring between 21 × 30 cm) to huge (typical billboard size, measuring greater than 2.0 × 2.5 m).

The classification of advertising density around schools was based on two categories: areas identified as hotspots and those designated as non-hotspot regions (Fig. 1). Hotspot classification was determined by the quantity of advertisements located schools, utilizing criteria established in earlier studies. In hotspot analysis, the previous study used the optimized hotspot analysis tool using Getis-Ord Gi* statistics to identify areas with a significantly higher density of adverts (at a 95 % significance level) [8].

The previous study also examined school data from 978 governmental and private schools in Semarang city. This data was received from the municipal education office on 15 May 2019, and may be accessed at <http://disdik.semarangkota.go.id>. The data contained school names and levels (primary, junior high, and senior high), along with addresses that we transformed into geocodes using Google Sheets and geocoding add-ons. Additional information regarding the methodologies and findings has been published in another source [8].

As a follow-up to the previous study [8], this research conducted a survey and interviews among high school students to assess smoking behaviors in smoking initiation. Smoking initiation is defined as an event indicator for being a never-smoker to an ever-smoker [22].

This study conducted a sample calculation using data from the Indonesian Global Youth Tobacco Survey (GYTS) and used Lameshow sample calculation [23].

$$n = \frac{\{(z_{1-\alpha})^2 p(1-p)\}}{d^2}$$

The data GYTS indicated that 50 % of young individuals reported being exposed to cigarette advertisements or promotions. Additionally, a margin of error of 5 % was considered. Based on these parameters, the minimum required sample size was 384 students, so this study got 400 samples [21].

A random selection procedure was employed to choose 20 high schools as the study setting, with 20 students interviewed at each school. Ethical Clearance was obtained from Semarang State University (242/KEPK/EC/2019). This study got permission from the school and consent from the students.

The eligibility criteria for student inclusion encompassed being male, this study specifically targeted males due to their significantly greater frequency of smoking in Indonesia. The respondents were at least ten years old, having spent at least one year in the school, and expressed willingness to participate in the study. The study employed adapted and validated versions of the GYTS questionnaires in Indonesian. Data collection was carried out by eleven trained enumerators over the period spanning September to December 2019.

This research used the Tripod (Transparent Reporting of a Multi-variable Prediction Model for Individual Prognosis or Diagnosis) Checklist to present findings on predicting youth smokers. The Tripod Statement aims to enhance the transparency of reporting in studies developing prediction models, particularly in the context of Youth's smoking behaviors [24,25].

Data set

The dataset in this study based on the previous study [21], consists of 400 data compromised five features: outdoor advertising (hotspot, non-hotspot); family factors's smoking status (no one, father is a smoker,

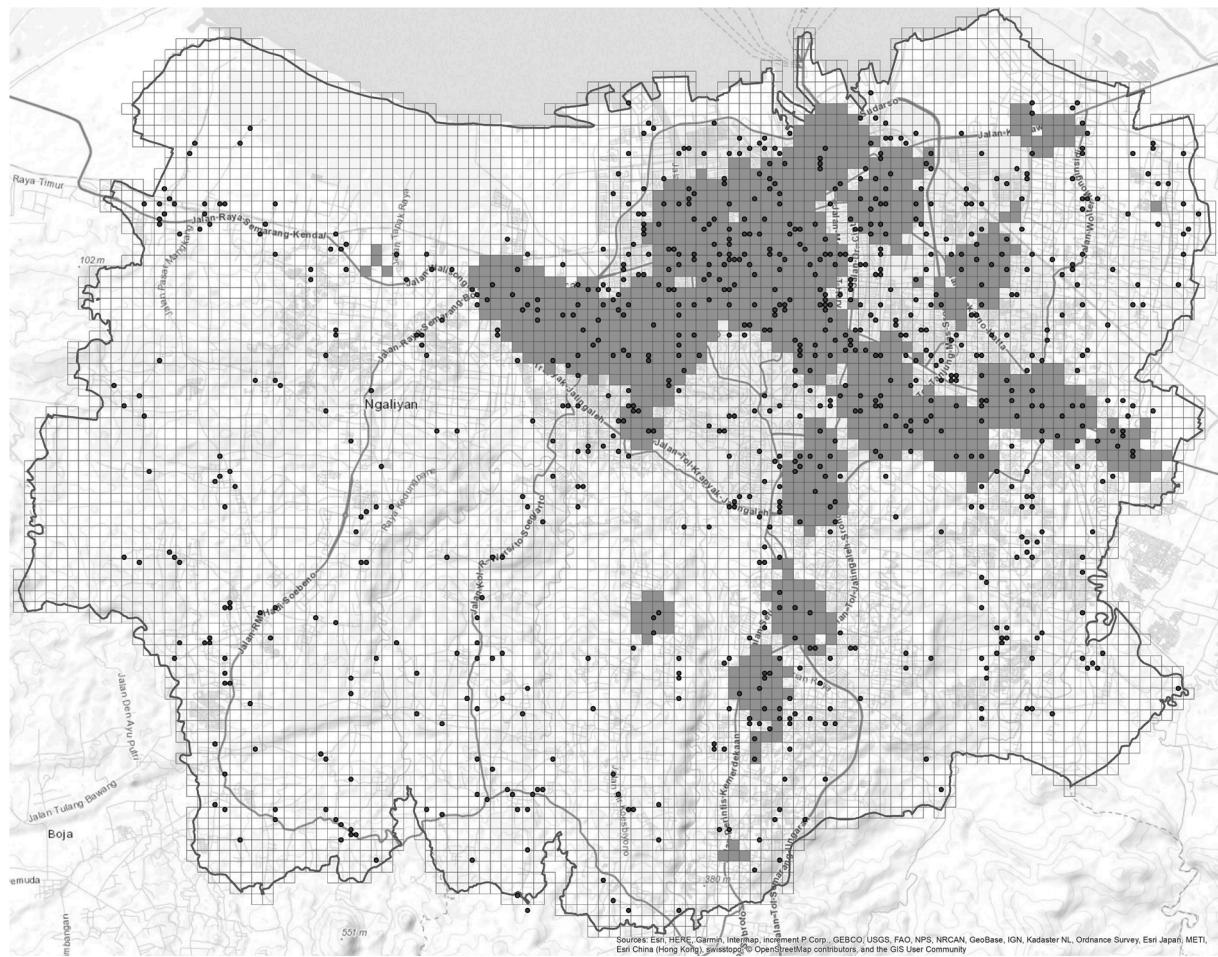


Fig. 1. The density of advertisements classifies as Hotspot and Non-Hotspot area.

Note: locations with a substantially (95 %) higher density of advertisements are represented by grey cells, while black dots indicate schools. [8].

mother is a smoker, both of them are smokers); peer smoker (no one, few, most of the peer are smokers, all are smokers); start of smoking (not smoker at, ≤ 7 years old, 8–9 years old, 10–11 years old, 12–13 years old, 14–15 years old); and Youth is smoker (yes, no). This data set is available at the Dryad repository.

Data analysis

Data presentation

The data set will be presented in descriptive analysis with cross-tabulation between respondent characteristics, outdoor advertising, peer and family factors, and smoking initiation. This study also represents the predictive model of smoking initiation among outdoor advertising, peer, and family factors with an accuracy performance model.

The performance of prediction models is typically assessed using accuracy, an evaluation metric commonly employed for classification tasks. Accuracy measures the proportion of correct predictions made by the model. Highly performance models result in precise and reliable predictions. A machine learning model performs well when it achieves an accuracy of over 70 %. A precision metric ranging from 70 % to 90 % is practical and in line with established norms in the industry [26].

Data mining

This study used the data set [27] to apply a data mining approach with a decision tree method, using Rapid Miner Ver. 9.8. Decision tree algorithms are usually used for variable selection, assessing the importance of variables, handling missing values, prediction, and data manipulation [28]. This study used a decision tree to assess the relative

importance of variables between outdoor advertising, parent's smoking status, and peer smoking with the smoking initiation.

The Decision Tree Classification Algorithm is like a typical tree structure, including roots, branches, and leaves. A decision tree is a tree where each node shows a feature (attribute), each link (branch) shows a decision (rule), and each leaf shows an outcome (categorical or continuous value) [29]. The decision tree has various algorithms for splitting attributes such as CART, ID3, C4.5, and CHAID. Table 1 describes that splitting criteria such as Information Gain, Gain Ratio, and Gini Index can be used [30]. This study applied those splitting criteria to generate the best performance model.

Table 1
Splitting equation.

Metric	Equation
Information Gain	$\text{Information Gain} = I(p, n) = \left(\frac{-p}{p+n} \right) \log_2 \left(\frac{p}{p+n} \right) - \left(\frac{n}{n+p} \right) \log_2 \left(\frac{n}{p+n} \right)$
Gain Ratio	$\text{Gain Ratio} = I(p, n) - E(A)$ $I(p, n) = \text{Information before splitting}$ $E(A) = \text{Information after splitting}$
Gini Index	$\text{Gini Index } (G) = \left(\frac{1}{2n^2\mu} \right) \sum_{j=1}^m \sum_{k=1}^m n_{jk} y_j - y_k $

Results

Respondent characteristics

This study distributed a standardized GYTS questionnaire to 400 respondents in 20 schools in Semarang City. [Table 2](#) shows that respondents are distributed almost equally in hotspot and non-hotspot areas (45 % and 55 %).

[Table 2](#) shows that 65.5 % of Youth have smoked at least once, and 24.5 % started to smoke at the age of 12 or 13 years. Data reveals that 24.8 % of students ≤ 11 years old ever tried to smoke, and students at the age of 13 years, tend to smoke regularly. There is no substantial difference in the proportion of Youth who smoked among hotspot and non-hot spot areas.

Peer factors

The peer factors consist of having peers who are smokers and the perception of respondents related to smokers (both boys and girls) in the neighborhood. [Table 3](#) shows that only 18 % of respondents do not have friends as smokers, while 10.3 % have all friends who are smokers.

[Table 3](#) shows that 92.7 % of youth with smoker friends have initiated to smoke, while for Youth who do not have smoker friends, only 29.2 % ever tried to smoke. Furthermore, the proportion of Youth who initiate smoking is likely higher in Youth with more attractive perceptions of smokers than in Youth with less attractive perceptions (80 % in boys and 74.1 % in girls)

[Table 3](#) also describes that respondents likely have no difference in perception about smoker boys while perception about smoker girls likely has negative perceptions such as naughty (52.8 %), less attractive (49.0 %), and having fewer friends (48.0 %).

Family factors

This study reveals that 79.8 % of families who have talked about the negative impact of smoking behavior; they have 64.3 % of the Youth who initiated smoking. It is <6.1 % compared to the family who never

Table 2
Respondent's characteristics.

Parameters	N	%	Initiation of smoking			
			Yes		No	
			N	%	N	%
Outdoor Tobacco Advertising						
Hotspot	180	45.0	116	64.4	64	35.6
Non-hotspot	220	55.0	146	66.4	74	33.6
Age						
≤ 11 years old	7	1.8	4	57.1	3	42.9
12 years old	35	8.8	11	31.4	24	68.6
13 years old	71	17.8	36	50.7	35	49.3
14 years old	97	24.3	58	59.8	39	40.2
15 years old	54	13.5	43	79.6	11	20.4
16 years old	74	18.5	63	85.1	11	14.9
17 years old	44	11.0	35	79.5	9	20.5
Above 17 years old	18	4.5	12	66.7	6	33.3
Youth Grade						
Grade 1	108	27.0	59	54.6	49	45.4
Grade 2	171	42.8	124	72.5	47	27.5
Grade 3	121	30.3	79	65.3	42	34.7
Ever tried to smoke						
Yes	262	65.5	—	—	—	—
No	138	34.5	—	—	—	—
Tried to smoke						
Never	138	34.5	—	—	—	—
≤ 7 years old	20	5.0	—	—	—	—
8 or 9 years old	19	4.8	—	—	—	—
10 or 11 years old	60	15.0	—	—	—	—
12 or 13 years old	98	24.5	—	—	—	—
14 or 15 years old	65	16.2	—	—	—	—

Table 3

Peer factors related to smoking initiation of youth in semarang city, indonesia.

Parameters Peer Factor	f	%	Initiation of smoking			
			Yes		No	
			f	%	f	%
Having Peer Smoker						
None	72	18.0	21	29.2	51	70.8
Few	172	43.0	106	61.6	66	38.4
Most	115	28.8	97	84.3	18	15.7
All of them	41	10.3	38	92.7	3	15.7
Perception of Smoker Male Friends						
Having more friends	79	19.8	62	78.5	17	21.5
Having fewer friends	77	19.3	31	40.3	46	59.7
No different	244	61.0	169	69.3	75	30.7
Perception of Smoker Female Friends						
Having more friends	31	7.8	22	71.0	9	29.0
Having fewer friends	192	48.0	120	62.5	72	37.5
No different	177	44.3	120	67.8	57	32.2
Being a smoker makes boys looks						
More attractive	50	12.5	40	80.0	10	20.0
Less attractive	119	29.8	65	54.6	54	45.4
No different	231	57.8	157	68.0	74	32.0
Being a smoker makes girls looks						
More attractive	27	6.8	20	74.1	7	25.9
Less attractive	196	49.0	125	63.8	71	36.2
No different	177	44.3	117	66.1	60	33.9
Perception about smoker boys						
Unconfident	82	20.5	65	79.3	17	20.7
Fool	70	17.5	37	52.9	33	47.1
Naughty	197	49.3	114	57.9	83	42.1
Success	15	3.8	13	86.7	2	13.3
Smart	7	1.8	6	85.7	1	14.3
Cool	29	7.3	27	93.1	2	6.9
Perception about smoker girls						
Unconfident	29	7.3	21	72.4	8	27.6
Fool	139	34.8	89	64.0	50	36.0
Naughty	211	52.8	134	63.5	77	36.5
Success	5	1.3	5	100	0	0.0
Smart	4	1.0	3	75.0	1	25.0
Cool	12	3.0	10	83.3	2	16.7

talked about the effect of smoking behavior ([Table 4](#)).

This study describes the family factors influencing students to smoke, while of youth whose father is a smoker, 73.2 % start to smoke, while among Youth having a smoker mother, 100 % tried to smoke ([Table 4](#)).

Table 4

Family factors related to smoking initiation of youth in semarang city, indonesia.

Parameters Family Factor	f	%	Initiation of smoking			
			Yes		No	
			f	%	f	%
Parents Smoker						
None	148	37.0	78	52.7	70	47.3
Both of them	7	1.8	6	85.7	1	14.3
Father	231	57.8	169	73.2	62	26.8
Mother	3	0.8	3	100	0	0.0
Do not know	11	2.8	6	54.5	5	45.5
A smoker in the house (one week)						
0 days	139	34.8	66	74.3	73	25.7
1 - 2 days	74	18.5	55	74.3	19	25.7
3 - 4 days	44	11.0	29	65.9	15	34.1
5 - 6 days	24	6.0	14	58.3	10	41.7
7 days	119	29.8	98	82.4	21	17.6
A smoker in the neighborhood (one week)						
0 days	75	18.8	28	37.7	47	62.7
1 - 2 days	80	20.0	47	58.8	33	41.3
3 - 4 days	60	15.0	41	68.3	19	31.7
5 - 6 days	42	10.5	34	81.0	8	19.0
7 days	143	35.8	112	78.4	31	21.7
The family talked about the side effect of smoking behavior						
Yes	319	79.8	205	64.3	114	25.7
No	81	20.3	57	70.4	24	29.6

Family members' behavior was captured through daily smoking habits; among Youth who have family members who smoke every day, 82.4 % tried to smoke. Furthermore, among Youth who have smokers in the neighborhood, 78.4 % of Youth tried to smoke.

Description among outdoor advertising, peer, and family factors

Table 5 shows the cross-tabulation between factors of outdoor advertising, peer, and family factors to generate youth smoking initiation. The percentage of youth who initiate to smoke getting bigger starting at 10–11 years old.

Table 5 shows that youth with peer smokers (at least a few smokers) likely become smokers rather than non-peer smokers; furthermore, youth with smoking parents and living in the hotspot advertising area

initiate smoking early.

Data mining

This study simultaneously analyzed three variables or features in the data set [27]: parents smoking status, peer smokers, and tobacco advertising around the school. The pattern in the data is explored using data mining approaches called decision trees function. This study reveals that Youth who have a smoker father and live in the hotspot area tend to smoke earlier than their counterpart. It is found that the hotspot area has double the number of Youth who smoke (eight Youth) than non-hotspot (four Youth), which started smoking at seven years of age.

This study applied the decision tree technique to obtain the model-related features between parents and peer smokers and tobacco

Table 5

Cross-tabulation between outdoor advertising, peer and family factors with youth's initiation smoking in indonesia.

Family factors	Initiation of smoking (years old)	Hotspot				Non-Hotspot				Total	
		Having Peer Smokers				Total	Having Peer Smokers				
		No one	Few	Many	All		No one	Few	Many		
Father	≤ 7 years	1 (12.5)	2 (25.0)	4 (50.0)	1 (12.5)	8 (7.8)	0 (0.0)	1 (25.0)	1 (25.0)	2 (50.0)	4 (3.9)
	8–9 years	1 (3.3)	1 (3.3)	0 (0.0)	1 (3.3)	3 (2.9)	1 (10.0)	4 (40.0)	2 (20.0)	3 (30.0)	10 (9.8)
	10–11 years	1 (5.9)	8 (47.1)	4 (23.5)	4 (23.5)	17 (16.6)	0 (0.0)	7 (46.7)	6 (40.0)	2 (13.3)	15 (14.7)
	12–13 years	1 (4.0)	13 (52.0)	8 (32.0)	3 (12.0)	25 (24.5)	1 (2.6)	19 (50.0)	15 (39.5)	3 (7.9)	38 (37.2)
	14–15 years	0 (0.0)	5 (20.0)	17 (68.0)	3 (12.0)	25 (24.5)	0 (0.0)	4 (17.4)	13 (56.5)	6 (26.1)	23 (22.5)
	Not smoker	7 (29.2)	11 (45.8)	4 (16.7)	2 (8.3)	24 (23.5)	7 (17.9)	23 (59.0)	9 (23.1)	0 (0.0)	39 (38.2)
	Total	11 (10.8)	40 (39.2)	37 (36.3)	14 (13.7)	102 (100)	9 (7.0)	58 (45.0)	46 (35.7)	16 (12.4)	129 (100)
Mother	8–9 years	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (50)
	12–13 years	1 (100)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (50)
	Total	1 (100)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)	2 (100)	0 (0.0)	0 (0.0)	2 (100)
	Both parents are smokers	8–9 years	1 (100)	0 (0.0)	0 (0.0)	1 (16.6)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (100)
None of the parents are smokers	10–11 years	0 (0.0)	1 (50.0)	1 (50.0)	0 (0.0)	2 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	14–15 years	2 (100)	0 (0.0)	0 (0.0)	0 (0.0)	2 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Not smoker	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	1 (16.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Total	3 (50)	1 (16.7)	1 (16.7)	1 (16.7)	6 (100)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (100)
	≤ 7 years	0 (0.0)	2 (100)	0 (0.0)	2 (0.0)	2 (2.9)	0 (0.0)	2 (40.0)	1 (20.0)	2 (40.0)	5 (6.3)
	8–9 years	1 (50.0)	1 (50.0)	0 (0.0)	0 (0.0)	2 (2.9)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	1 (1.23)
	10–11 years	3 (33.3)	3 (33.3)	0 (0.0)	3 (33.3)	9 (13.0)	5 (29.4)	9 (52.9)	1 (5.9)	1 (11.8)	17 (21.5)
	12–13 years	2 (22.2)	5 (55.6)	0 (0.0)	2 (22.2)	9 (13.0)	2 (11.1)	9 (50.0)	4 (22.2)	3 (16.7)	18 (22.7)
	14–15 years	2 (22.2)	4 (44.4)	2 (22.2)	1 (11.1)	9 (13.0)	4 (66.7)	4 (33.3)	0 (0.0)	0 (0.0)	6 (7.6)
	Not smoker	4 (10.5)	11 (28.9)	1 (2.6)	22 (57.9)	38 (55.0)	3 (9.4)	17 (53.1)	0 (0.0)	12 (37.5)	32 (40.5)
Do not know if parents are smokers	Total	12 (17.4)	26 (37.7)	3 (4.3)	28 (40.6)	69 (100)	14 (17.7)	39 (49.4)	6 (7.6)	20 (25.3)	79 (100)
	≤ 7 years	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (11.1)
	8–9 years	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (11.1)
	12–13 years	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	1 (50)	2 (66.7)	0 (0.0)	0 (0.0)	1 (33.3)	3 (33.3)
	Not smoker	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	1 (50)	0 (0.0)	3 (75.0)	0 (0.0)	1 (25.0)	4 (44.4)
	Total	0 (0.0)	1 (50.0)	0 (0.0)	1 (50.0)	2 (100)	2 (22.2)	4 (44.4)	1 (11.1)	2 (22.2)	9 (100)

advertising around the school using four criteria: Information Gain, Gini Index, Gain Ratio, and Accuracy. Fig. 2 shows the decision tree resulting in two models; only the Accuracy criterion has a different model among the criteria. All the models have a performance classification of 77.5 %.

Fig. 2 shows the model produced by various weighting criteria. The evaluation with Gini Index, Information Gain, and Gain Ratio, shows that the resulting model is similar. However, the model produced by the Accuracy criterion metric has a slightly different model but gives the

same performance result as the other criterion.

Fig. 3 shows that Peer Smoker status with the branch value of All Peer Smoker has 92.3 % confidence value that youth start smoking. Similarly, *Peer Smokers with a branch value of Most of Peer Smokers has an 80.2 % confidence value*. Then, the branch value *Few of Peer Smoker* must be considered the *Parent Smoker* node. Both parent smokers and mother smokers have confidence values up to 100 %, except the branch value of Parent Smoker is *Do Not Know*. This branch flows to the next

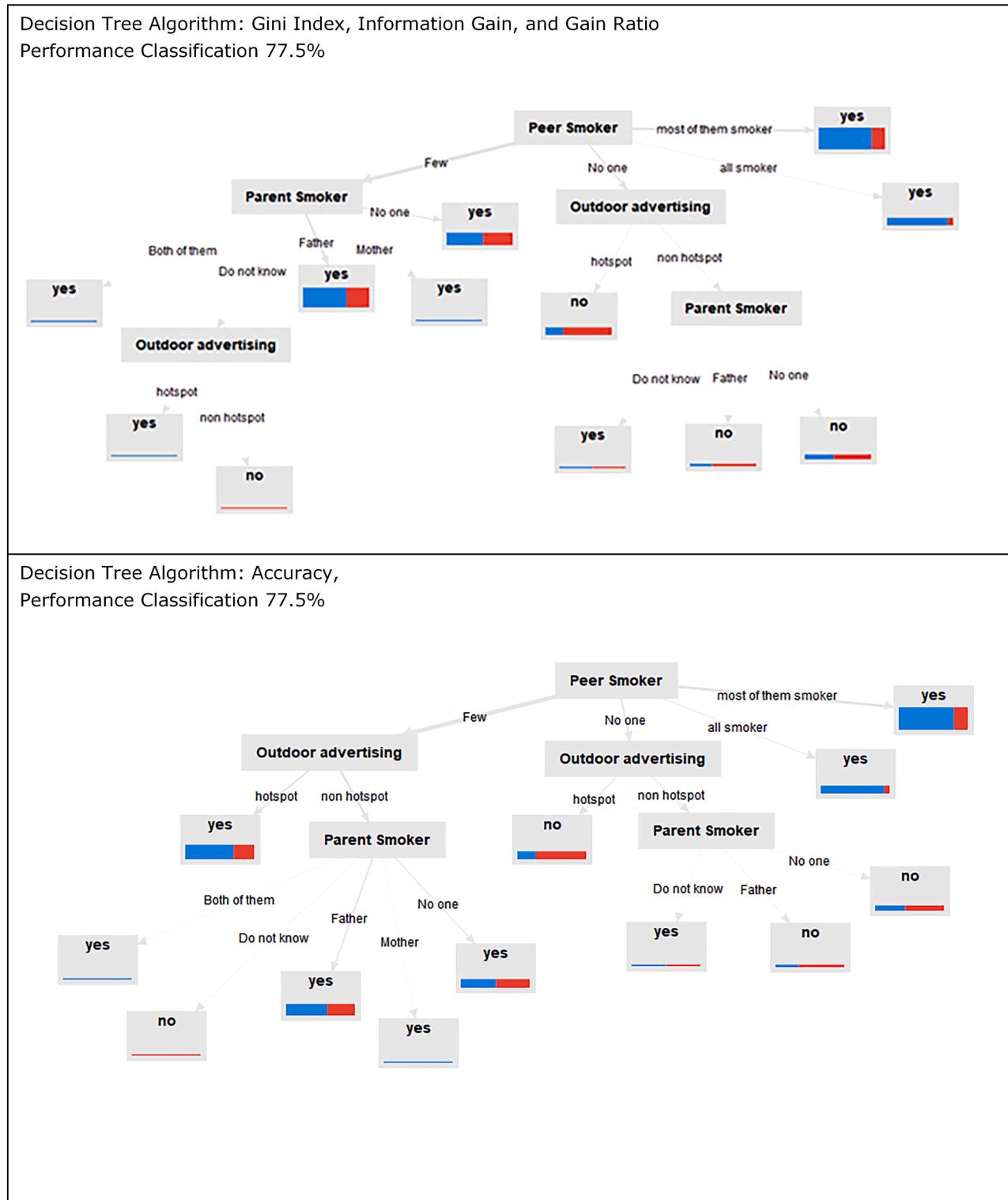


Fig. 2. Decision tree model of youth smokers.

Tree Description from Gini Index, Information Gain, Gain Ratio

Peer Smokers = Few

- | Parent Smokers = Both of them: yes {yes=2, no=0} (Confidence 100%)
- | Parent Smokers = Do not know
- | | Outdoor advertising = hotspot: yes {yes=1, no=0} (Confidence 100%)
- | | Outdoor advertising = non hotspot: no {yes=0, no=1} (Confidence 100%)
- | Parent Smokers = Father: yes {yes=48, no=25} (Confidence 65.8%)
- | Parent Smokers = Mother: yes {yes=1, no=0} (Confidence 100%)
- | Parent Smokers = No one: yes {yes=25, no=20} (Confidence 55.5%)

Peer Smokers = No one

- | Outdoor advertising = hotspot: no {yes=7, no=19} (Confidence 73%)
- | Outdoor advertising = non hotspot
- | | Parent Smokers = Do not know: yes {yes=1, no=1} (Confidence 50%)
- | | Parent Smokers = Father: no {yes=2, no=4} (Confidence 66.6%)
- | | Parent Smokers = No one: no {yes=7, no=9} (Confidence 56.2%)

Peer Smokers = All smokers

yes {yes=24, no=2} (Confidence 92.3%)

Peer Smokers = Most of them smokers; yes {yes=65, no=16} (Confidence 80.2%)

Fig. 3. Tree description from three weighting criteria.

consideration node, *Outdoor Advertising*. Its branch value *Hotspot* figures out that the youth start smoking and *non-Hotspot* shows that they would not start smoking with a confidence level of 100 %, respectively.

Additionally, the model indicates that the *peer smoker* is the root of the model, followed by the *Outdoor Advertising* node in case the youths do not have peer smokers. The findings demonstrate that the *Hotspot* value within *Outdoor Advertising* strongly discourages the initiation of smoking, with a confidence level of 73 % (Fig. 3). Furthermore, the *non-hotspot* value of *Outdoor Advertising* must be considered the *Parent Smoker* node, where the value of the *Father* and *No One* of the parent smokes predict that the Youth would not start smoking. Furthermore, *Do Not Know* the Value of *Parent Smoker* denoted that Youth would start smoking despite the confidence level of these three branches' value being <70 %.

Discussion

Various prior studies deployed machine learning techniques to explore the complexity of smoking behavior in adults. [31-33] Maryam et al. emphasize that data mining techniques are promising in predicting smoking behavior. This study explores the factors influencing Youth to smoke, such as peer smoking, outdoor tobacco advertisements, and whether the parents are smokers. In our research, it was observed that 63.5 % of the subjects had experimented with smoking, and 24.5 % had initiated smoking at the ages of 12 or 13. These outcomes closely parallel those documented in studies conducted in the United Kingdom. Our investigation highlighted a heightened susceptibility to smoking among Youth aged between 11.7 and 12.7 years when subjected to both influences from smoking siblings and peer interactions [34]. The age at which one first began smoking cigarettes every day is the most significant explanatory factor for the prediction of the average number of cigarettes used per day, followed by educational attainment and psychological distress, according to a prior study [31].

However, our study revealed that a higher number of advertisements contributed to increased smoking Youth. Additionally, a separate study identified the normative social image as a risk factor for Youth when it comes to smoking behavior [35].

These findings show the importance of making a policy on reducing the exposure of outdoor tobacco advertisements, as a previous study also demonstrated an increased density of 70 % and 51 % within a 100-meter radius of primary schools [8]. Creating a safe environment for Youth is crucial to prevent peer pressure from influencing new smokers.

The first factor related to the Youth's decision to smoke in this study

is the peer factor because it lies in the root node of the decision tree. In case the Youth do not have peer smokers, outdoor advertising is the second factor affecting the Youth's smoking behavior. Youths with few peer smokers, the second layer factor is parents. Studies showed that youths 8 years old became addicted to tobacco because of peer pressure [36]. Another study showed that peer pressure significantly influenced cigarette smoking [37]. Peer pressure positively affects cigarette smoking [38,39]. Furthermore, the Accuracy method generated a different model in Youth who have few peer smokers; the outdoor advertising lies in the second layer, unlike the other methods. The previous study described that a youth's smoking habit is most closely connected to that of the best friend and friends in general, with a minor relationship to parental smoking behavior. Nevertheless, longitudinal outcomes showed a considerably reduced effect for the best friend and friends in general in predicting smoking behavior one year later. [40]

The family's smoking status also impacts smoking among Youth [39]. This study found that Youth with smoking fathers are more likely to start smoking earlier, while for Youth whose mothers are smokers, the chance is 100 % (confidence 100 %). The previous study found that Youth with smoking parents have tried smoking compared to those with non-smoking parents [41]. The acceptance of Youth related to the smoking behavior of adults because of the perception that smoking is one of the coping strategies of the parents; Youth understand that adults have problems [42].

The previous study found that Youth of mothers who smoked throughout pregnancy had a higher risk of developing asthma and chronic wheezing [43]; furthermore, tobacco use is a significant asthma risk factor [44]. The role of the family in practicing a healthy lifestyle influences the Youth's behavior, including tobacco use [45]. The intervention strategy needs to focus on the family and broader environment to prevent Youth from smoking and healthy Youth.

This study used data mining as a novel approach to understand human behavior instead of a generic statistical approach [46]. Data mining is increasingly popular and essential in the health community such as health care and public health [47,48]. It reveals previously unknown patterns and trends in data and uses that information to build predictive models such as treatment outcomes, and predictive behavior [33,49]. Various functions of data mining can be applied to acquire new knowledge in health care [50]. Moreover, recent studies have used data mining to develop questionnaires and understand health behavior [51, 52].

Different classification algorithms are available in the literature on

the data mining method. However, the decision tree is the most used method because of its ease of execution and easier understanding than other classification algorithms. Frequently used algorithms to develop decision trees such as CART, C4.5, CHAID, QUEST, and ID3 are available. The decision tree methodology is commonly used to establish classification systems based on multiple covariates or develop prediction algorithms for a target variable [28].

The model resulting from the decision tree pattern in Fig. 2 shows that the root node is the peer factor in both models, which is the key to the initiation of smoking in the Youth. Simultaneously, parents are the main factor in developing Youth's smoking behavior if the Youth has peer smokers. A previous study described that even though parents are smokers, if the Youth are exposed to antismoking programs or awareness programs about the negative impact of smoking, they are less likely to try cigarettes [53].

Our findings are consistent with an earlier study that found significant associations between smoking among youth (but not among adults) with density and proximity to outdoor advertising in Semarang City. Youth at schools with a medium and high density of outdoor advertising (OTA) were up to 2.16 times more likely to smoke than those with low-density [21].

Our study has at least two limitations. First, our study was conducted in an urban setting and should not stand for the whole country. Second, this study only explored the initiation of the Youth who tried to smoke, not examining whether the Youth were active smokers or not.

Conclusion

The study findings underscore that peer influence stands as the leading determinant in encouraging Youth to start smoking, followed by the impact of tobacco advertising and parental smoking behaviors. Moreover, this investigation indicates that non-hotspot regions within outdoor advertising do not significantly influence Youth to initiate smoking. Conversely, the smoking habits of mothers have a notable contribution to shaping whether Youth engage in smoking behaviors or not. These outcomes bear significant implications for policy formulation not only within Indonesia but also at the global level.

Further study

Further research should also examine rural districts to explore regional and socioeconomic variations. Second, our research only interviewed high school students because data showed that smoking initiation is penetrating in the younger age groups in Indonesia; further studies should also assess smoking behavior among primary school students.

Informed consent

"All procedures followed were as per the ethical standard of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration Of 1975, as revised in 2000. Informed consent was obtained from all the respondents to be included in the study."

Data

Data available at <https://doi.org/10.5061/dryad.76hdr7sxg>

Funding

Research Funding support was provided by the Center for Islamic Economics and Business, Universitas Indonesia, with funding awarded by Bloomberg Philanthropies to Johns Hopkins University. Its content is solely the authors' responsibility and does not necessarily represent the official views of Bloomberg Philanthropies or Johns Hopkins University.

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of State University of Semarang (Number: 242/KEPK/EC/2019 Date: 11 October 2019)

CRediT authorship contribution statement

Enny Rachmani: Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Sri Handayani:** Validation, Project administration, Conceptualization. **Kriswiharsi Kun Saptorini:** Project administration, Investigation, Data curation. **Nurjanah:** Writing – review & editing, Visualization, Methodology. **Dian Kusuma:** Writing – review & editing, Visualization. **Abdillah Ahsan:** Writing – review & editing, Project administration, Conceptualization. **Edi Jaya Kusuma:** Writing – review & editing, Visualization, Data curation. **Suleman Atique:** Writing – review & editing, Validation. **Jumanto Jumanto:** Writing – review & editing.

Declaration of competing interest

The authors declare that we have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgments

Thanks to the Center for Islamic Economics and Business, Universitas Indonesia, for giving us the opportunity to conduct this research, and the Faculty of Health Science, Universitas Dian Nuswantoro, for all the support to the authors. Thanks to Prof. Ir. Zainal Arifin Hasibuan MLS., . Ph.D., for suggestions in the manuscript.

References

- [1] M.o.H.o. Indonesia, *Health Profile of Indonesia 2022*, Ministry of Health, Jakarta, 2022.
- [2] WHO, *Factsheet 2018: Indonesia NCD and Tobacco Surveillance*, WHO, 2018.
- [3] M.O. Robichaud, et al., How media stories in low- and middle-income countries (LMICs) Discussed the US Food and Drug Administration's (FDA's) Modified Risk Tobacco Product (MRTP) Order for IQOS, *Nicotin Tob. Res.* (2023).
- [4] C. Reynolds, Tobacco advertising in Indonesia: "the defining characteristics for success, *Tob. Control* 8 (1) (1999) 85–88.
- [5] P.A.S. Astuti, B. Freeman, It is merely a paper tiger." Battle for increased tobacco advertising regulation in Indonesia: content analysis of news articles, *BMJ Open* 7 (9) (2017) e016975.
- [6] Ministry of Health, I., WHO, and CDC, *Global Youth Tobacco Survey Fact Sheet Indonesia*. 2019. p. 2.
- [7] Megatsari, H., et al., *Visibility and hotspots of outdoor tobacco advertisement around educational facilities without an advertising ban: geospatial analysis in Surabaya City, Indonesia*. 2019. 5.
- [8] N. Nurjanah, et al., Density and hotspots of outdoor tobacco advertising around schools in Semarang, Indonesia: geospatial analysis, *Int. J. Tuberc. Lung Dis.* 24 (7) (2020) 674–680.
- [9] Ministry of Health, I., *Indonesia Fact Sheet: WHO SEARO. Global Youth Tobacco Survey 2014 dan 2019* 2019.
- [10] M. Nicther, et al., Reading culture from tobacco advertisements in Indonesia, *Tob. Control* 18 (2) (2009) 98–107.
- [11] O.G. Chido-Amajoyi, et al., Association of tobacco advertising, promotion and sponsorship (TAPS) exposure and cigarette use among Nigerian adolescents: implications for current practices, products and policies, *BMJ Global Health* 2 (3) (2017).
- [12] C. Alexander, et al., Peers, schools, and adolescent cigarette smoking, *J. Adolesc. Health* 29 (1) (2001) 22–30.
- [13] L. Huntington-Moskos, et al., The peer and family smoking index: a valid measure of secondhand smoke exposure in adolescents, *J. Adolesc. Health* 58 (4) (2016) 446–450.
- [14] A. Bandura, *Social foundations of thought and action*, ed. E. Cliffs, Vol. 1986 (1986) 2.
- [15] L. Mercken, et al., Dynamics of adolescent friendship networks and smoking behavior, *J. Social. Networks* 32 (1) (2010) 72–81.
- [16] S.C. Montgomery, et al., Peer social network processes and adolescent health behaviors: a systematic review, *J. Preventive Med.* 130 (2020) 105900.

- [17] J. Alves, et al., The role of parental smoking on adolescent smoking and its social patterning: a cross-sectional survey in six European cities, *J. J. Public Health Res.* 39 (2) (2017) 339–346.
- [18] J. Leonardi-Bee, M.L. Jere, J. Britton, Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis, *J. Thorax* 66 (10) (2011) 847–855.
- [19] E.G. García, et al., Estudio de los factores asociados al uso de sustancias en la adolescencia mediante Reglas de Asociación, *Adicciones* 22 (4) (2010) 293–300.
- [20] U. Prüss, et al., Patterns of behaviour of juvenile smokers and non-smokers, *Prax. Kinderpsychol. Kinderpsychiatr.* 53 (5) (2004) 305–318.
- [21] S. Handayani, et al., Is youth smoking related to the density and proximity of outdoor tobacco advertising near schools? Evidence from Indonesia, *Int. J. Environ. Res. Public Health* 18 (5) (2021) 2556.
- [22] J.M. Nonnemacher, M.C. Farrelly, Smoking initiation among youth: the role of cigarette excise taxes and prices by race/ethnicity and gender, *J. Health Econ.* 30 (3) (2011) 560–567.
- [23] S.K. Lwanga, S. Lemeshow, W.H. Organization, *Sample Size Determination in Health studies: a Practical Manual*, World Health Organization, 1991.
- [24] K.G. Moons, et al., Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration, *J. Annals of Int. Med.* 162 (1) (2015) W1–W73.
- [25] Collins, G.S., et al., *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement*. 2015. 131(2): p. 211–219.
- [26] Barkved, K. *How To Know if Your Machine Learning Model Has Good Performance*. 2022 [cited 2023 June, 10, 2023]; Available from: <https://www.obviously.ai/post/machine-learning-model-performance>.
- [27] E. Rachmani, Sri Handayani, K.K. Saptorini, *Why do youths initiate to smoke? A Data mining analysis on Tobacco Advertising, Peer, and Family Factors for Indonesian Youths*, U.D. Nuswantoro, Editor, Dryad, 2021.
- [28] Y.-Y. Song, L. Ying, Decision tree methods: applications for classification and prediction, *J. Shanghai Archives of Psychiatry* 27 (2) (2015) 130.
- [29] S.D. Jadhav, H. Channe, Efficient recommendation system using decision tree classifier and collaborative filtering, *Int. Res. J. Eng. Technol.* 3 (8) (2016) 2113–2118.
- [30] H.H. Patel, P. Prajapati, Study and analysis of decision tree based classification algorithms, *Int. J. Comput. Sci. Eng.* 6 (10) (2018) 74–78.
- [31] S.S. Moon, et al., Decision tree models for characterizing smoking patterns of older adults, *Expert Syst. Appl.* 39 (1) (2012) 445–451.
- [32] M. Abo-Tabik, et al., Decision Tree Model of Smoking Behaviour. 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE, 2019.
- [33] Xu, Y., et al., *Prediction of Smoking Behavior From Single Nucleotide Polymorphisms With Machine Learning Approaches*. 2020. 11: p. 416.
- [34] A. Swan, R. Creeser, M. Murray, When and why children first start to smoke, *Int. J. Epidemiol.* 19 (2) (1990) 323–330.
- [35] Andrews, J.A., et al., *The effect of subjective normative social images of smokers on children's intentions to smoke*. 2008. 10(4): p. 589–597.
- [36] M. Senthilingam, Chain-smoking children: indonesia's ongoing tobacco epidemic, *CNN News* 9 (2017). Tersedia di, <https://edition.cnn.com>. Diakses pada.
- [37] C.T. Leshargie, et al., The impact of peer pressure on cigarette smoking among high school and university students in Ethiopia: a systemic review and meta-analysis, *PLoS One* 14 (10) (2019) e0222572.
- [38] N. Dereje, S. Abazinab, A. Girma, Prevalence and predictors of cigarette smoking among adolescents of Ethiopia: school based cross sectional survey, *J. Child Adolescen Behavior* 3 (182) (2014) 2.
- [39] X. Wei, et al., Study on smoking status in families and its influence on children in Shandong province, Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi 36 (8) (2015) 815–820.
- [40] H. De Vries, et al., Parents' and friends' smoking status as predictors of smoking onset: findings from six European countries, *J. Health Education Res.* 18 (5) (2003) 627–636.
- [41] Loke, A.Y. and Y.P.I.J.J.o.A.N. Wong, *Smoking among young children in Hong Kong: influence of parental smoking*. 2010. 66(12): p. 2659–2670.
- [42] Milton, B.S., et al., 'My Mum and Dad said it Calms You Down': children's Perceptions of Smoking as a Coping Strategy. 2012. 26(2): p. 89–99.
- [43] Holst, G.J., et al., Air pollution and family related determinants of asthma onset and persistent wheezing in children: nationwide case-control study. 2020. 370.
- [44] Hollams, E.M., et al., Persistent effects of maternal smoking during pregnancy on lung function and asthma in adolescents. 2014. 189(4): p. 401–407.
- [45] M. Stojanovic-Tasic, et al., Influence of family smoking habits and passive smoking on smoking status among Belgrade university students, *Subst. Use Misuse* 51 (3) (2016) 310–317.
- [46] Hand, D.J., *Data Mining: new Challenges for Statisticians*. 2000. 18(4): p. 442–449.
- [47] T. E., et al., Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance, *J. Am. Med. Inform. Assoc.* 5 (4) (1998) p. Aug 1998.
- [48] N. Lavrac, et al., Data mining and visualization for decision support and modeling of public health-care resources, *J. Biomed. Inform.* 40 (4) (2007) 438–447.
- [49] S. Hansun, Prediction case notification rates for tuberculosis in eight countries, *J. Hunan Univ. Natural Sci.* 47 (9) (2020).
- [50] Hian Chye Koh, G. Tan, Data Mining Applications in Healthcare, *J. Healthc. Inf. Manag.* 19 (2) (2011). Vol. 19, No. 2.
- [51] D. Spruijt-Metz, et al., Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research, *J. Translat. Behav. Med.* 5 (3) (2015) 335–346.
- [52] E. Rachmani, et al., Mining Medication behavior of the completion leprosy's multi-drug therapy in Indonesia, in: 2018 International Seminar on Application for Technology of Information and Communication, IEEE, 2018.
- [53] C. Jackson, D. Dickinson, a. medicine, Enabling parents who smoke to prevent their children from initiating smoking: results from a 3-year intervention evaluation, *J. Archiv. Pediatrics* 160 (1) (2006) 56–62.