ORIGINAL ARTICLE

Classification of Type 2 Diabetes Mellitus Using Machine Learning in Sleman District of Yogyakarta Special Region, Indonesia

Dhite Bayu Nugroho[1,2], Hemi Sinorita[1], Bowo Pramono[1], Robikhul Ikhsan[1], Vinayanti Susanti[1], Mawaddah Ar Rochmah[3]

[1] Department of Internal Medicine, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, 55281
[2] Clinical Epidemiology and Biostatistics Unit, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia 55281
[3] Department of Neurology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia, 55281

ABSTRACT

Introduction: Diabetes mellitus type 2 (T2DM) is a chronic-multifactorial disease with a high disease burden. Identifying the risk factors for type 2 diabetes mellitus can help in designing prevention strategies. Surveillance system data can be utilized to accurately predict the prevalence of diseases in a community using machine learning algorithm. The aim of this study was to determine the performance of machine learning and to identify important features in classifying T2DM in the Health and Demographic Surveillance System (HDSS) population of the Sleman region of Yogyakarta, Indonesia. **Methods:** The first two cycles of the Sleman HDSS database were obtained, and factors such as demographics, risky foods, diet composition, and comorbidity were evaluated. After pre-processing the data, we employed binary classification of T2DM using logistic regression and the random forest method. The performance of the two models was then compared, and the imported features were reported. **Results:** There were 4,611 subjects included in this study including 463 with self-reported T2DM. Significant differences, such as age, level of education, monthly food expenditure, consumption of coffee or other caffeinated beverages, intake of herbs and instant noodles, and health issues, such as hypertension and stroke, were identified between the T2DM and non-DM groups. Apart from fat, rice is the most predominant food in all types of diet compositions. **Conclusion:** The random forest machine-learning algorithm shows superior performance, with hypertension being the most important feature in the classification of a self-reported T2DM in the Sleman HDSS population.

Keywords: Diabetes mellitus type 2, Sleman HDSS, Machine learning, Important features

Corresponding Author:
Dhite Bayu Nugroho, PhD
Email: dhite@ugm.ac.id
Tel: +62 274 545458

INTRODUCTION

Diabetes mellitus type 2 is a chronic multifactorial disease caused by environmental and genetic factors (1). Globally, the prevalence of diabetes mellitus is increasing, particularly in Asian countries, which accounts for more than 60% of the world's population with diabetes (2). Diabetes prevalence continues to rise in Indonesia, and, according to a report by the International Diabetes Federation (IDF) in 2009, there were approximately 7.3 million persons living with T2DM in Indonesia, growing to 10.3 million in 2017 and projected to reach 16.7 million by 2045 (3). According to the 2018 Basic Health Research Data (RISKESDAS), the number of T2DM patients in Indonesia increased from 6.9 percent in 2013 to 10.9 percent in 2018. The growth in patient numbers happened primarily in large cities, one of which was the Special Region Yogyakarta (DIY), which has the third-highest number of patients with diabetes in the country after Jakarta and East Kalimantan (4).

Since 2015, the Sleman Health and Demographic Surveillance System (HDSS) has been monitoring the health and demographic conditions of the residents of the Sleman Regency in Yogyakarta Province, Indonesia. Sleman HDSS regularly collects data on demographics, health problems, and health services. Demographic information is gathered on migration, births, deaths, causes of death, and socioeconomic status. Data are also collected on infectious and non-communicable diseases, maternal and child health as well as the use of and access to healthcare services (5).

Big data is often described as having elements such as volume, velocity, variety, and truthfulness. Big data and its associated analytic tools hold the promise of vastly improved health promotion and disease prevention strategies. They allow a more precise identification of at-risk populations through a more comprehensive understanding of human health and disease, including the interaction between genetic, lifestyle, and environmental determinants of health (6). In addition, big data can help to better understand and address modifiable behavioral risk factors that contribute to a large portion of the non-communicable disease burden. A classification model that is accurate in determining the type 2 diabetes mellitus status can be used to identify risk variables that contribute significantly to the prevalence of type 2 diabetes mellitus and can be used to develop disease prevention strategies. Machine learning (ML) technology is a machine designed to learn on its own without human intervention. Machine learning is built on other fields such as statistics, mathematics, and data mining in order for machines can learn by evaluating data without having to be programmed (7). Machine learning can retrieve existing data using its own commands. ML may also analyze existing data as well as new data to perform specific tasks. In this study, supervised machine learning techniques were utilized to analyze data on the features of the sleman HDSS population, and this information was then used to predict the status of self-reported type 2 diabetes mellitus. There is currently no model that uses machine-learning approaches to classify the Type 2 Diabetes Mellitus group in the Sleman population.

The main objective of this study is to compare two machine-learning models capable of predict self-reported diabetes mellitus type 2 in the Sleman HDSS population, namely random forest and logistic regression. Another goal of this study is to determine the prevalence of type 2 diabetes in this community.

MATERIALS AND METHODS

Data source and sample design

This study is a descriptive observational study with secondary analysis based on data from the Sleman HDSS, and the data collection methods, briefly described in this study, partially replicate their wording (5). Since 2015, the Sleman HDSS database has been updated annually to reflect the annual cycle of data collection. In brief, the Sleman HDSS collected data annually from longitudinal surveys which were administered to subjects sampled using two-stage cluster sampling designs and using Census Blocks (CBs) in 17 Sleman sub-districts. 216 CBs were randomly selected from the 3,513 CBs, resulting in 184 CBs in urban areas and 32 CBs in rural areas. Twenty-five households were randomly selected from each CB using systematic random sampling. Since the Sleman HDSS uses households as the sampling unit and an annually updated questionnaire, the longitudinal survey's response rate and the completeness of the data variables varied across cycles. However, the Sleman HDSS is a reliable source of data for public health and demographic research. This is due to its high response rate of more than 97 percent from more than 4,942 households, which is the study's calculated minimum sample size, having more than 19,724 respondents in the study's first two cycles in 2015 and 2016 (5).

The Sleman HDSS used INDEPTH Network to develop standardized questionnaires for each cycle, as did other HDSSs in Indonesia. Pre-coded questions about non-communicable diseases, such as stroke, hypertension, diabetes mellitus, and cardiovascular disease, as well as questions about family meals, were collected during home visits by trained enumerators, using Computer-Assisted Personal Interviewing (CAPI). Non-communicable diseases were recorded as self-reported diagnoses, and family meals from the previous seven days were recalled (5).

Food composition in the family meals of diabetes mellitus type 2 patients in the Sleman District, Yogyakarta

During each interview, the participants recalled the principal ingredients of their family meals. The main ingredients recorded corresponded to the family meals prepared in the preceding seven days. Each day, respondents estimated the quantities of primary ingredients in the prepared family meal menu and then totaled them across the seven days. The primary ingredient quantities were recorded in grams.

For the purposes of the study, the recorded amounts of key ingredients were adapted for daily family meals." According to the Indonesian Food Composition Table, the calorie, protein, fat, carbohydrate, and fiber contents of each primary ingredient were then computed as a single food (8). Due to the fact that the data recorded the primary ingredients prior to the cooking procedure, the nutritional values reflected an estimate of the original analytical values. Mixed or composite foods that were not made by the family were eliminated from this study due to the lack of knowledge regarding the quantity of the primary ingredients prior to processing or cooking.

Statistical analysis

The characteristics of T2DM patients in the rural and urban areas of the Sleman District were determined using descriptive analyses. The mean+SD of the numerical data was calculated using an independent sample t-test. Categorical data were expressed as proportions or ratios (in percentages) and analyzed using the chi-square test. All statistical analyses were conducted in RStudio using the meta-packages "tidyverse" and "tidymodels" (9). The Institutional Review Board of the Faculty of Medicine, Public Health, and Nursing, at Universitas Gadjah Mada, approved this study under the number KE/0473/04/2021.

Exploratory data analysis and data processing

We used data from Sleman HDSS cycles 1 and 2, which had been collected in 2015 and 2016. We performed data cleaning, using the "tidy" principle, where each column represents a research variable and each row represents the result of observation (10). We only included data in the analysis for which there were no "non-available" observations. We then created a dummy variable from all nominal variables except the outcome variable, self-reported diabetes mellitus type 2. Because the number of subjects with type 2 diabetes was not equal to the number of non-DM type 2 subjects in this classification study, we used the Synthetic Minority Oversampling Technique (SMOTE) to balance the numbers between the two groups before proceeding to machine-learning analysis.

Machine-learning algorithm

We used a machine-learning algorithm to identify individuals in the Sleman HDSS population who self-reported having T2DM. In this study, two types of machine learning were used: logistic regression and random forest. After pre-processing the data, we evaluated the performance of these two classification models by comparing their Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), sensitivity, specificity, and accuracy. Finally, we have reported the top ten variables with the greatest importance for the best model. The schematic methodology of machine learning analysis is presented in Figure 1.

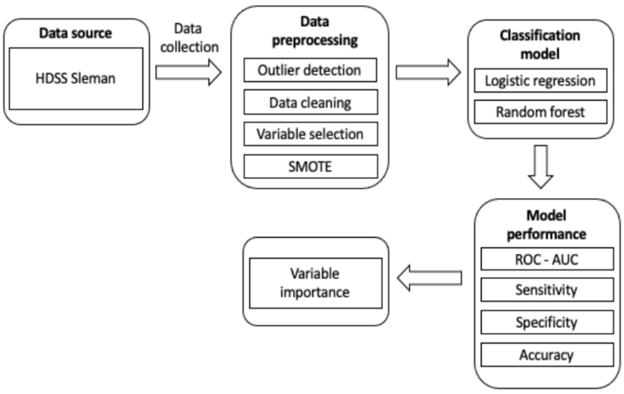


Figure 1: Methodology for machine-learning analysis

RESULTS

Characteristics of diabetes mellitus type 2 patients in rural and urban areas of Sleman Regency, Yogyakarta, Indonesia

We combined the first two cycles of the Sleman HDSS, which included data on family meals and non-communicable diseases, including DM type 2. The number of subjects taking part with complete data for all study parameters, was 4611, 463 of whom self-reported having T2DM. Table I, which summarizes

Table I: Baseline characteristics

	Diabetes mellitus type 2			
	Overall (n = 4611)	No (n = 4148)	Yes (n = 463)	p-value
Sex = male (%)	1658 (36.0)	1474 (35.5)	184 (39.7)	0.082
Region = urban (%)	3207 (69.6)	2853 (68.8)	354 (76.5)	0.001*
Age (mean (SD))	48.64 (13.59)	48.14 (13.53)	53.16 (13.33)	<0.001*
Marital status				0.053
Divorced	589 (12.8)	542 (13.1)	47 (10.2)	
Married	3801 (82.4)	3415 (82.3)	386 (83.4)	
Not yet married	221 (4.8)	191 (4.6)	30 (6.5)	
Education (%)				<0.001*
Elementary school	1039 (22.5)	930 (22.4)	109 (23.5)	
Junior high school	787 (17.1)	724 (17.5)	63 (13.6)	
High school	1736 (37.6)	1567 (37.8)	169 (36.5)	
College or above	758 (16.4)	653 (15.7)	105 (22.7)	
No formal education	291 (6.3)	274 (6.6)	17 (3.7)	
Occupation (%)				<0.001*
Farmer	356 (7.7)	334 (8.1)	22 (4.8)	
House wife	1367 (29.6)	1225 (29.5)	142 (30.7)	
Labourer	759 (16.5)	725 (17.5)	34 (7.3)	
Private employee	400 (8.7)	367 (8.8)	33 (7.1)	
Self-employed	733 (15.9)	655 (15.8)	78 (16.8)	
Unemployed	238 (5.2)	206 (5.0)	32 (6.9)	
Others	758 (16.4)	636 (15.3)	122 (26.3)	
Insurance = yes (%)	3031 (65.7)	2715 (65.5)	316 (68.3)	0.250
Hypertension = yes (%)	1528 (33.1)	1273 (30.7)	255 (55.1)	<0.001*
Stroke = yes (%)	63 (1.4)	46 (1.1)	17 (3.7)	<0.001*
Cancer = yes (%)	64 (1.4)	59 (1.4)	5 (1.1)	0.698
Food expenses (mean (SD))	869.71 (647.67)	850.83 (623.06)	1038.88 (818.11)	<0.001*
High-risk foods				
Sweet Food / Drink (Mean (SD))	1.86 (1.85)	1.87 (1.20)	1.76 (4.60)	0.230
Salty Food (Mean (SD))	1.70 (3.68)	1.68 (3.55)	1.87 (4.66)	0.305
Fatty/Fried Food (Mean (SD))	2.15 (3.07)	2.13 (2.84)	2.31 (4.62)	0.225
Roasted Food (Mean (SD))	1.27 (3.73)	1.20 (2.91)	1.83 (7.89)	0.001*
Processed Meat / Chicken / Food With Preservative Seasonings (Mean (SD))	0.95 (3.48)	0.92 (2.95)	1.22 (6.50)	0.080
Coffee (Mean (SD))	1.66 (2.35)	1.63 (1.23)	1.87 (6.45)	0.040*
Caffeinated Drinks Not Coffee (Mean (SD))	0.93 (2.44)	0.90 (1.39)	1.16 (6.48)	0.033*
Salted Fish (Mean (SD))	1.52 (2.43)	1.49 (1.38)	1.72 (6.46)	0.054
Herb (Mean (SD))	1.43 (3.99)	1.38 (3.28)	1.90 (7.89)	0.008*
Instant Noodles (Mean (SD))	0.94 (2.34)	0.90 (1.19)	1.30 (6.47)	0.001*

*Statistically significant

the descriptive characteristics of the patients with type 2 DM, shows that there is no statistically significant difference in gender or marital status between patients with T2DM and non-patients with T2DM. Most of the people with type 2 diabetes live in urban areas, are older, and have a college education or higher. Non-DM type 2 participants are more likely to work as farmers or laborers, whereas T2DM participants are more likely to be unemployed. There is no difference between the two groups in terms of health insurance ownership.

Among the non-communicable diseases that had comorbidity characteristics, stroke and hypertension occurred more frequently in the T2DM group, whereas cancer occurred equally in both groups.

Food composition in family meals of diabetes mellitus type 2 patients in Sleman District, Yogyakarta
The mean daily intake of high-risk diets was reported. The T2DM group consumed significantly more roasted food, coffee, caffeinated beverages, herbs, and instant noodles than the non-DM type 2 group. Other risky foods, such as sweet/salty foods, fatty/fried foods, processed meat/chicken/food with preservative seasonings, and salted fish were also consumed more frequently by the T2DM group, although the difference was not statistically significant when compared to the non-T2DM group.

Additionally, using a lollipop chart in Figure 2, we have presented descriptive data on food composition. Apart from fat, rice is the most predominant food in all types of diet compositions, accounting for between 40% and 80% of the total. Notably, the non-DM type 2 group consumed a higher percentage of diet energy, fiber, fat, and the protein derived from rice, than the type 2 DM group. Similarly, protein derived from tempeh, energy, and carbohydrates derived from sugar represented a higher percentage in the non-DM type 2 group, even though the total percentage was less than 20%. The type 2 diabetes group consumed a greater proportion of fatty oils, energy oils, chicken protein, and tuber fiber.

Machine learning analysis
The receiver operator characteristics curves for each model are shown in Figure 3. Our results show that the logistic regression model does slightly better than the random forest model in classifying self-reported T2DM (Table II).

Based on the findings of the performance analysis model, random forest performs better than logistic regression, as

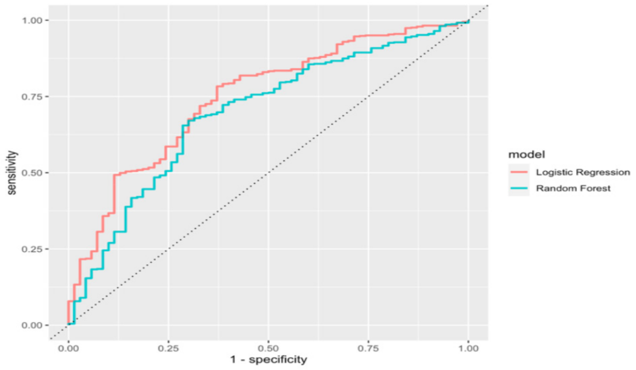
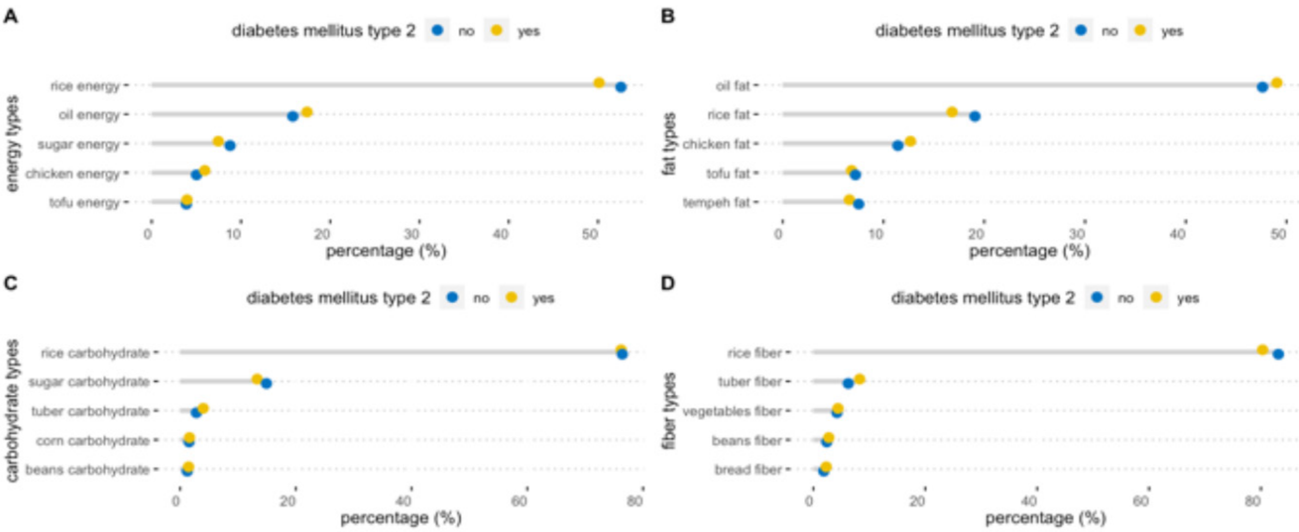


Figure 3: ROC curve of logistic regression and the random forest method



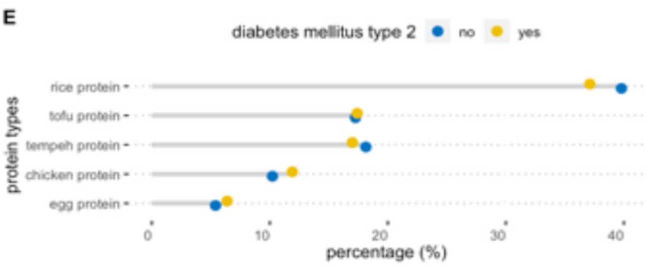


Figure 2: Food composition percentage between diabetes mellitus type 2 and non-between diabetes mellitus type 2

Table II: Model performance

	Accuracy	Roc_Auc	Sensitivity	Specificity
Logistic regression	0.677	0.687	0.687	0.579
Random forest	0.894	0.667	0.990	0.032

evidenced by a higher rate of accuracy. Therefore, we will only look at the feature importance of the random forest technique.

Visualization of feature importance, in Figure 4, shows that hypertension status had the highest influence on self-reported T2DM in the HDSS cohort. The consumption of high-risk foods, such as sweetened foods or beverages, living in a city, having health insurance, and having a high school degree, can all be included in the random forest algorithm to determine if a subject has self-reported T2DM.

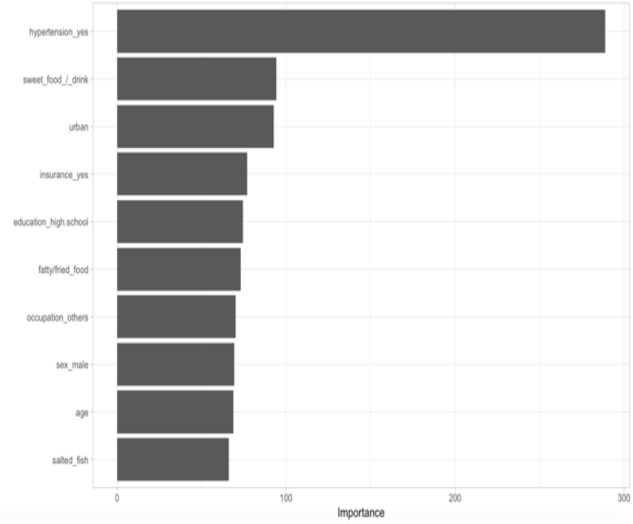


Figure 4: Feature importance based on the random forest algorithm

DISCUSSION

Machine learning is gaining popularity for classifying non-communicable diseases, such as stroke (11), coronary heart disease (12), and diabetes (13), both in clinical and population investigations (14). This is consistent with the increasing availability of massive amounts of data in a short time and in various formats, which is referred to as big data analysis (15).

The two algorithms used in this study are both commonly used approaches for categorizing binary outcomes, such as self-reported diabetes mellitus type 2. Logistic regression is a well-established classification technique. It is a generalization of conventional regression in that it may model a binary variable representing the probability of an event occurring or not occurring (16). Additionally, a random forest is a form of ensemble classifier that classifies data by utilizing numerous data tables (DTs). Following that, either the classification with the greatest "votes" (for discrete classification results) or the forest average, is chosen by the forest (for numeric classification outcomes). Since the RF technique evaluates the outcomes of several DTs, it can help reduce the variance introduced by a single DT in a given dataset (16).

Consistent with previous research, established risk factors for diabetes mellitus, such as hypertension (17), consumption of sugary foods or beverages (18), and increasing age (19), all play a role in identifying persons with self-reported T2DM in the Sleman HDSS cohort. As a result of the demographic surveillance research conducted in other countries, such as Myanmar (20), Ethiopia (21), Senegal (22), and Peru (23), Sleman's population demonstrates that persons residing in urban areas have a higher risk of developing non-communicable diseases, particularly T2DM. One of the factors is the rapid rate of urbanization or migration from rural to urban regions (24), as living in urban areas is associated with a sedentary lifestyle, unhealthy eating habits, and high-stress levels, all of which pose a risk of developing type 2 diabetes mellitus (25, 26).

There were 12 occupations recorded in this survey, but only the top six were identified in the report, whereas the others were classified as other occupations. One of the other occupations constitutes one of the most relevant factors in the Sleman HDSS population's classification of T2DM. Yogyakarta, which has the greatest percentage (34%) of retirees compared to other occupation groups (data not shown), is known as a retirement city, with many Indonesian residents from various regions of the country choosing to settle there.

Type 2 diabetes mellitus is associated with chronic complications, such as coronary heart disease, retinopathy, and neuropathy, which require long-term therapy (2). A large proportion of people with type two diabetes mellitus have health insurance (27). As a result, insurance ownership ranks fourth on the list of the most important variables.

Interestingly, diet composition, a significant predictor of type 2 diabetes mellitus in many studies, was not a significant predictor in the machine-learning random forest method. However, variables in diet composition were important using the machine-learning logistic regression method to discriminate against those with type 2 diabetes mellitus.

The study's limitation lies mainly with the validation of the subjects' diabetes mellitus status, which was verified solely from personal accounts or reports from the head of the household at the time this survey was conducted. Additionally, the ability to perform disease classification using these two methods of machine learning is relatively limited, necessitating the collection of additional data for potential explanatory variables.

CONCLUSION

The random forest machine-learning algorithm performs best for the categorization of self-reported T2DM in the Sleman HDSS population, with hypertension being the most significant feature. This machine learning model helps identify individuals in the Sleman HDSS population with type 2 diabetes mellitus who could be the focus of management strategies for type 2 diabetes mellitus.

AKNOWLEDGEMENTS

This publication used secondary data provided by the Sleman Health and Demographic Surveillance System (HDSS).

REFERENCES

1. Goyal, R., & Jialal, I. (2022). Diabetes Mellitus Type 2. In StatPearls. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK513253/
2. Ramachandran, A., Snehalatha, C., Shetty, A. S., & Nanditha, A. (2012). Trends in prevalence of diabetes in Asian countries. World Journal of Diabetes, 3(6), 110–117. https://doi.org/10.4239/wjd.v3.i6.110
3. Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B. B., Stein, C., Basit, A., Chan, J. C. N., Mbanya, J. C., Pavkov, M. E., Ramachandaran, A., Wild, S. H., James, S., Herman, W. H., Zhang, P., Bommer, C., Kuo, S., Boyko, E. J., & Magliano, D. J. (2022). IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Research and Clinical Practice, 183, 109119. https://doi.org/10.1016/J.DIABRES.2021.109119
4. Kemenkes RI. Riskesdas 2018. Development. 2018:1-220.
5. Dewi, F. S. T., Lestari, S. K., Niamilah, I., & Wulandari, H. (2020). Developing a system to utilize a surveillance data for evidence-based public health interventions: Sleman HDSS's experience. Journal of Community Empowerment for Health, 3(1), 8–17. https://doi.org/10.22146/JCOEMPH.47144
6. Gall C, Suzuki E. 5 Big data: A new dawn for public. OECD Health Policy Studies Health in the 21st Century Putting Data to Work for Stronger Health Systems: Putting Data to Work for Stronger Health Systems. 2019 Nov 21:146. Available from: https://www.oecd-ilibrary.org/sites/f24cb567-en/index.html?itemId=/content/component/f24cb567-en.
7. Kementerian Kesehatan Republik Indonesia. Pedoman metode melengkapi nilai gizi bahan makanan pada tabel komposisi pangan Indonesia (imputated and borrowed values) [Internet]. Jakarta: Kementerian Kesehatan Republik Indonesia; 2017 [cited 2020 January 5). Available from: https://www. panganku.org/assets/files/publikasi/351bb3
8. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2016. Available from: https://www.R-project.org/
9. Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), 1–23. https://doi.org/10.18637/JSS.V059.I10
10. Lee, H., Lee, E.-J., Ham, S., Lee, H.-B., Lee, J. S., Kwon, S. U., Kim, J. S., Kim, N., & Kang, D.-W. (2020). Machine Learning Approach to Identify Stroke Within 4.5 Hours. Stroke, 51(3), 860–866. https://doi.org/10.1161/STROKEAHA.119.027611
11. Motwani, M., Dey, D., Berman, D. S., Germano, G., Achenbach, S., Al-Mallah, M. H., Andreini, D., Budoff, M. J., Cademartiri, F., Callister, T. Q., Chang, H. J., Chinnaiyan, K., Chow, B. J. W., Cury, R. C., Delago, A., Gomez, M., Gransar, H., Hadamitzky, M., Hausleiter, J., … Slomka, P. J. (2017). Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. European Heart Journal, 38(7), 500–507. https://doi.org/10.1093/EURHEARTJ/EHW188
12. Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Medical Informatics and Decision Making, 19(1). https://doi.org/10.1186/S12911-019-0918-5
13. Doupe, P., Faghmous, J., & Basu, S. (2019). Machine Learning for Health Services Researchers. Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research, 22(7), 808–815. https://doi.org/10.1016/j.jval.2019.02.012
14. Borges do Nascimento, I. J., Marcolino, M. S., Abdulazeem, H. M., Weerasekara, I., Azzopardi-Muscat, N., Gonзalves, M. A., & Novillo-Ortiz, D. (2021). Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies. Journal of Medical Internet Research, 23(4), e27275. https://doi.org/10.2196/27275
15. Couronnй, R., Probst, P., & Boulesteix, A. L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics, 19(1). https://doi.org/10.1186/S12859-018-2264-5
16. Akalu, Y., & Belsti, Y. (2020). Hypertension and Its Associated Factors Among Type 2 Diabetes Mellitus Patients at Debre Tabor General Hospital, Northwest Ethiopia. Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy, 13, 1621. https://doi.org/10.2147/DMSO.S254537
17. Stanhope, K. L. (2016). Sugar consumption, metabolic disease and obesity: The state of the controversy. Critical Reviews in Clinical Laboratory Sciences, 53(1), 52–67. https://doi.org/10.3109/10408363.2015.1084990

18. Charvat, H., Goto, A., Goto, M., Inoue, M., Heianza, Y., Arase, Y., Sone, H., Nakagami, T., Song, X., Qiao, Q., Tuomilehto, J., Tsugane, S., Noda, M., & Inoue, M. (2015). Impact of population aging on trends in diabetes prevalence: A meta-regression analysis of 160,000 Japanese adults. Journal of Diabetes Investigation, 6(5), 533–542. https://doi.org/10.1111/JDI.12333
19. Aung, W. P., Htet, A. S., Bjertness, E., Stigum, H., Chongsuvivatwong, V., & Kjшllesdal, M. K. R. (2018). Urban-rural differences in the prevalence of diabetes mellitus among 25-74 year-old adults of the Yangon Region, Myanmar: two cross-sectional studies. BMJ Open, 8(3). https://doi.org/10.1136/BMJOPEN-2017-020406
20. Gebremedhin, G., Enqueselassie, F., Deyessa, N., & Yifter, H. (2020). Urban-Rural Differences in the Trends of Type 1 and Type 2 Diabetes Among Adults Who Received Medical Treatment from Public Hospitals in Resource-Poor Community Tigray, Ethiopia. Diabetes, Metabolic Syndrome and Obesity : Targets and Therapy, 13, 859–868. https://doi.org/10.2147/DMSO.S238275
21. Seck, S. M., Dia, D. G., Doupa, D., Diop-Dia, A., Thiam, I., Ndong, M., & Gueye, L. (2015). Diabetes Burden in Urban and Rural Senegalese Populations: A Cross-Sectional Study in 2012. International Journal of Endocrinology, 2015. https://doi.org/10.1155/2015/163641
22. Seclen, S. N., Rosas, M. E., Arias, A. J., & Medina, C. A. (2017). Elevated incidence rates of diabetes in Peru: report from PERUDIAB, a national urban population-based longitudinal study. BMJ Open Diabetes Research & Care, 5(1). https://doi.org/10.1136/BMJDRC-2017-000401
23. Gassasse, Z., Smith, D., Finer, S., & Gallo, V. (2017). Association between urbanisation and type 2 diabetes: an ecological study. BMJ Global Health, 2(4). https://doi.org/10.1136/BMJGH-2017-000473
24. Compernolle, S., de Cocker, K., Teixeira, P. J., Oppert, J. M., Roda, C., Mackenbach, J. D., Lakerveld, J., McKee, M., Glonti, K., Rutter, H., Bardos, H., Cardon, G., & de Bourdeaudhuij, I. (2016). The associations between domain-specific sedentary behaviours and dietary habits in European adults: a cross-sectional analysis of the SPOTLIGHT survey. BMC Public Health, 16(1), 1–9. https://doi.org/10.1186/S12889-016-3708-3
25. Mendenhall, E., Shivashankar, R., Tandon, N., Ali, M. K., Venkat Narayan, K. M., & Prabhakaran, D. (2012). Stress and Diabetes in Socioeconomic Context: A Qualitative Study of Urban Indians. Social Science & Medicine (1982), 75(12), 2522. https://doi.org/10.1016/J.SOCSCIMED.2012.09.040
26. You, C. H., Choi, J. H., Kang, S., Oh, E. H., & Kwon, Y. D. (2018). Association between supplementary private health insurance and visits to physician offices versus hospital outpatient departments among adults with diabetes in the universal public insurance system. PLoS ONE, 13(4). https://doi.org/10.1371/JOURNAL.PONE.0192205