Journal Sensi
Print ISSN: 2461-1409
Online ISSN: 2655-5298

Research Article Open Access (CC–BY-SA)

Visual Analysis Of Body Signals In Smoker Data To Understand Health Impacts On Python

Anas Nasrulloh[*1], Muhamad Yusuf [2], Ibnu Mas'ud[3], Tubagus Toifur[4], Aolia Ikhwanudin[5], Agianto Syamhalim[6]
[1,2,3,4,5,6]Information Systems Study Program, Faculty of Computer Science, Institut Teknologi Tangerang Selatan, Tangerang
e-mail: anas@itts.ac.id[*1] , yusuf@itts.ac.id[2], ibnu@itts.ac.id[3], tubagus@itts.ac.id[4], aikwanudin@itts.ac.id[5] , agianto@itts.ac.id[6]

Abstract

In this study, researchers found that 44.4% of people had blood pressure less than 120/80, 58.2% had fasting blood sugar <= 99, 47.7% had hemoglobin > 17.2, and 62% had oral problems and teeth caused by smoking. This research was carried out by visualizing the impact of smoking on blood pressure, fasting blood sugar, hemoglobin and the mouth and teeth in the body using body signal of smoking data obtained from Kaggle which was analyzed and presented in the form of a pie chart using Python.

Keywords—Smoker, Tobacco, Body signals; Visualization, Blood pressure, Blood sugar.

1. INTRODUCTION

The Ministry of Health released the results of a global survey on tobacco use in adults (Global Adult Tobacco Survey – GATS) which was carried out in 2011 and repeated in 2021 involving 9,156 respondents. In his findings, over the last 10 years, there has been a significant increase in the number of adult smokers by 8.8 million people, namely from 60.3 million in 2011 to 69.1 million smokers in 2021. "This is a challenge for all of us to make efforts to stop smoking," said Deputy Minister of Health Dante Saksono Harbuwono at the Launch of Global Survey Data on Tobacco Use in Indonesian Society in 2021 (GATS 2021) which coincided with the commemoration of World No Tobacco Day, Tuesday (31/5) in Jakarta. The GATS survey results also show an increase in the prevalence of e-smoking up to 10 times, from 0.3% (2011) to 3% (2021). Regarding warning labels on cigarette packages, survey results show that the number of exposure to health warnings has increased from 77.2% (2011) to 77.6% (2021). Another finding is that cigarettes have a big impact on the socio-economic impact of society.

Currently, cigarettes are the second largest expenditure for poor people, higher than spending on nutritious food. Then the desire to stop smoking is quite high, namely 63.4% and 43.8% are trying to stop smoking. The Ministry of Health is currently continuing to open Quitline services for citizens who need smoking cessation counseling services. This needs to be improved considering that only 38.9% visit health services to stop smoking [19]. The visualization carried out is changing rigid tabular data into graphs, diagrams, etc. which can show changes and differences in data more clearly. As a result, the data display resulting from regular analysis becomes more beautiful with good visuals using matplotlib and pandas. From the test results, it is clear that the most significant spread of COVID-19 is in Indonesia, where since the 65th day it has surged [15]. Data visualization also shows that Twitter users who discuss higher education are more focused on campus life and new student admissions. This research evaluates the stemming stage because the results show that using stemming will eliminate semantics between words and visualize semantic relationships in tweets related to higher education [16]. This research was carried out in three stages, namely retrieving, preprocessing, and then visualization. The retrieving process is used to retrieve Tweet data from the BMKG account on Twitter. The preprocessing stage functions to obtain the results of the analysis of the quality of earthquake information. After that, the results of the data analysis on timeliness, relevance, completeness of the data, and accuracy of the data are made to conclusions stating the quality of information on the spread of earthquakes, then the results are displayed in the form of graphic visualization [17]. Research using unstructured data from social media which is parsed using web scraping will be visualized in Google Data Studio so that it is easier to analyze and represent to Kediri city government executives. Web scraping is carried out using the Python programming language and data extraction using Excel. The desired research result is the presentation of insight data in the form of tables and graphs on each social media account belonging to the Kediri City Government, such as quantitative likes, shares, and comments from followers [18].

To carry out data analysis, researchers here use a dataset sourced from www.kaggle.com which amounts to 55,692 data and researchers explore the data on the data, then visualize it and present it in the form of a pie chart for data about body signals from smoking such as blood pressure, sugar blood, hemoglobin, oral and tartar. Results from visualization for the largest percentage in EL1R1, blood pressure was 44.4% in TS4R4, blood sugar was 58.2% in FBS1, hemoglobin was 47.7% in MH3, oral and tartar was 62% in OYTY, eyesight was 48.5% in EL1R1, and triglyceride and LDL was 41.1% in AT1L1.

2. RESEARCH METHODOLOGY

The purpose of this study is for visualization of body signals from Python-based smoking to make it easier to understand the body's signals for smoking. To find out, exploring data must be done. This is done so that we can find out the percentage value of each visualization of the body's signal.

Researchers explore data on the data and then visualize and present it in the form of pie charts for data about body signals from smoking such as eyesight, blood pressure, blood sugar,

hemoglobin, oral, and tartar. The result of visualization for the largest percentage of eyesight is 48.5% on EL1R1, blood pressure of 44.4% on TS4R4, blood sugar is 58.2% on FBS1, hemoglobin by 47.7% in MH3, oral and tartar by 62% on OYTY.

Data in the study amounted to 55,692 data and data came from the website with the address www.kaggle.com. This dataset is a collection of basic health biological signal data.
The goal is to determine the presence or absence of smoking through bio-signals.

The following information is related to the dataset used:

Name of dataset: body_signal _of_smoking
About Dataset: This dataset is a collection of basic health biological signal data. The goal is to find out the presence or absence of cigarettes through bio-signal.
Dataset contents: ID: Index, Gender, Age: 5-Years Gap, Height (CM), Weight (Kg), Waist (CM): Waist Circumference Length, Eyesight (Left), Eyesight (Right), Hearing (Left), Hearing (Right), Systemic: Blood Pressure, Relaxation: Blood Pressure, Fasting Blood Sugar, Cholesterol: Total, Triglyceride, HDL: Cholesterol Type, LDL: Cholesterol Type, Hemoglobin, Urine Protein, Serum Creatinine, AST: , Alt: Glutamic Oxaloacetic Transaminase Type, GTP: γ-GTP, Oral: Oral Examination Status, Dental Caries, Tartar: Tartar Status, Smoking

In building a model there are 3 main stages, namely data understanding, data preparation, and data modeling. Each stage can be seen in Figure 3.1 below:

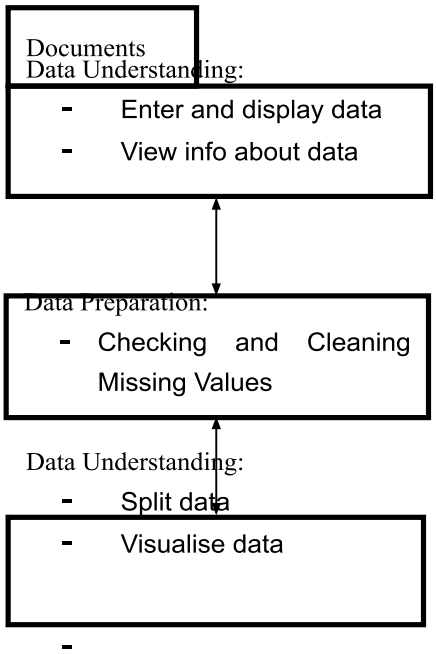

Figure 3.1. Development of research models

Data Understanding:

• Enter and Display

Data Researchers enter data and then display it in a form that is easy to read and understand, such as a table.

• Data Exploration

Researchers analyze data to look for patterns, trends, or anomalies; to gain a deeper understanding of what the data is saying.

Data Preparation:

• Checking and Cleaning Missing Values

Researchers examine the data for missing or empty values and then take action to correct them. For example, if there is incomplete data, they can fill in the missing values with reasonable estimates, delete rows or columns with many missing values, or use other methods to handle missing data so that the data analysis remains accurate and reliable.

Data Modeling:

• Split Data

Researchers divide the data they have into two or more parts. The data is divided into training data and testing data. The training data is used to build the model, while the testing data is used to test how well the model works with new data that has never been seen before.

• Data Visualization

Researchers visualize data to see patterns, trends, and relationships in data more easily.

3. RESULTS AND DISCUSSION

The relationship between smoking and blood pressure

Researchers split data and then present the relationship between smoking and blood pressure in the form of pie charts and provide labels such as TS1R1, TS2R2, TS3R3, and TS4R4. Visualization can be seen in the following Figure 4.1:

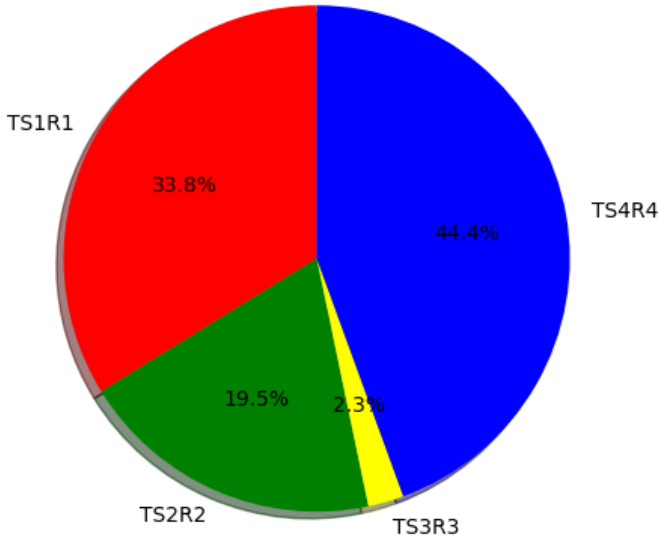

Figure 4.1. The percentage of the relationship between smoking and blood pressure

From Figure 4.1 above, it can be explained as follows:

- TS1R1 means smoking, there is a systolic response> 120 and relaxation> 80, has a percentage of 33.8%,
- TS2R2 means smoking, there is a systolic response> 120 and relaxation <80, has a percentage of 19.5%,
- TS3R3 means smoking, there is a systolic response <120 and relaxation> 80, has a percentage of 2.3% which is the smallest percentage,
- TS4R4 means smoking, there is a systolic response <120 and relaxation <80, has a percentage of 44.4% which is the largest percentage.

The relationship between smoking and blood sugar

Researchers split data and then present the relationship between smoking and blood sugar in the form of pie charts and provide labels such as FBS1, FBS2, and FBS3. Visualization can be seen in the following Figure 4.2:

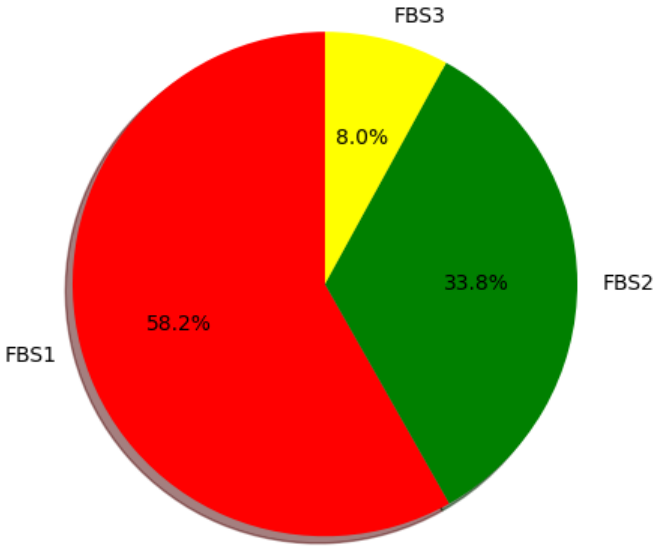

Figure 4.2. The percentage of the relationship between smoking and blood sugar

From Figure 4.2 above, it can be explained as follows:

- FBS1 means smoking, there is a Fasting Blood Sugar response <= 99, has a percentage of 58.2% which is the largest percentage,
- FBS2 means smoking, there is a Fasting Blood Sugar response> = 100 and <= 125, has a percentage of 33.8%,
- FBS3 means smoking, there is a Fasting Blood Sugar response> = 126, has a percentage of 8% which is the smallest percentage.

The relationship between smoking and hemoglobin

Researchers split data and then present the relationship between smoking and hemoglobin in the form of pie charts and provide labels such as FH1, FH2, FH3, MH1, MH2, and MH3.
Visualization can be seen in Figure 4.3 below:

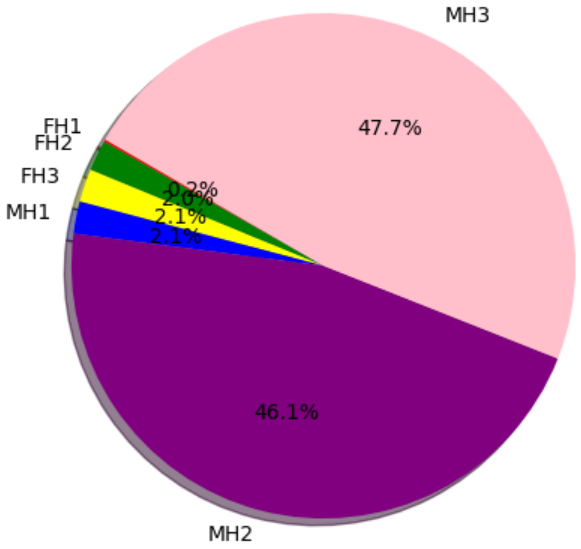

Figure 4.3. The percentage of the relationship between smoking and hemoglobin

From Figure 4.3 above, it can be explained as follows:

- FH1 means smoking, there is a hemoglobin response <12.1, has a percentage of 0.2%, which is the smallest percentage.
- FH2 means smoking, there is a hemoglobin response> = 12.1 and <= 15.1, has a percentage of 2.0%,
- FH3 means smoking, there is a hemoglobin response> 15.1, has a percentage of 2.1%
- MH1 means smoking, there is a hemoglobin response <13.8, has a percentage of 2.1%,
- MH2 means smoking, there is a response of hemoglobin> = 13.8 and <= 17.2, has a percentage of 46.1%,
- MH3 means smoking, there is a hemoglobin response> 17.2, and has a percentage of 47.7% which is the largest percentage.

The relationship between smoking and oral and tartar

Researchers split data and then present the relationship between smoking and oral and tartar in the form of pie charts and provide labels such as OYTY, OYTN, ONTY, and ONTN. Visualization can be seen in Figure 4.4 below:

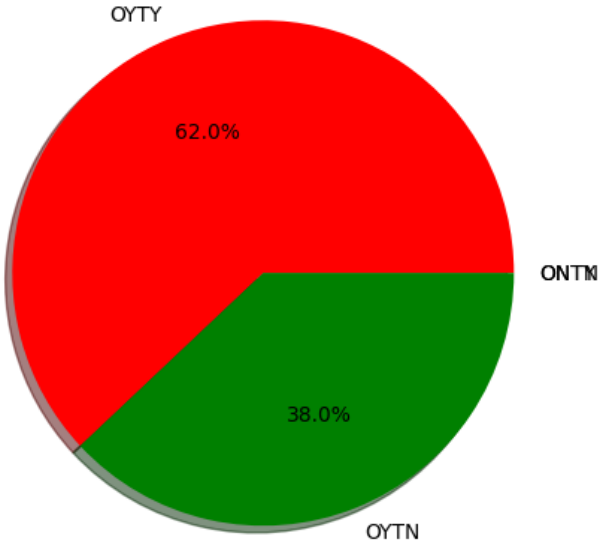

Figure 4.4. The percentage of the relationship between smoking and oral and tartar

From Figure 4.4 above, it can be explained as follows:

- OYTY means smoking, there is an oral response and tartar response, which has a percentage of 62% which is the largest percentage.
- OYTN means smoking, there is an oral response and no tartar response, has a percentage of 38%
- ONTY means smoking, there is no oral response and there is a tartar response, which has a percentage of 0% which is the smallest percentage.
- ONTN means smoking, there is no oral response and no tartar response, has a percentage of 0% which is the smallest percentage.

The relationship between smoking and eyesight

Researchers split data and then present the relationship between smoking and eyesight in the form of pie charts and provide labels such as EL1R1, EL2R2, EL3R3, and EL4R4. Visualization can be seen in Figure 4.5 below:

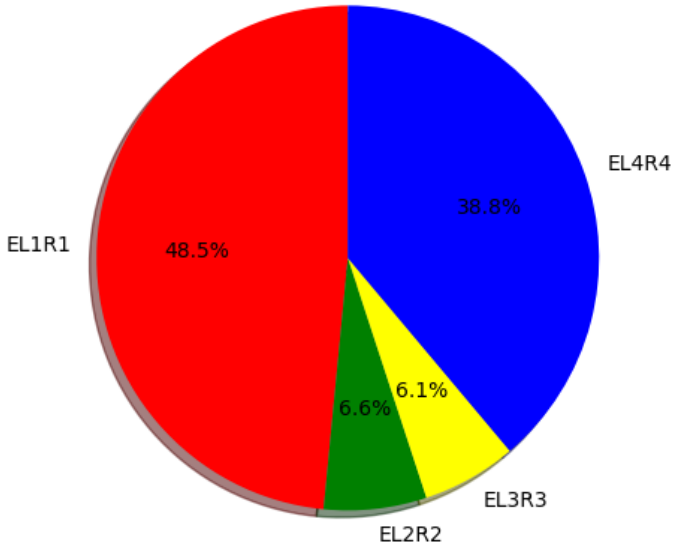

Figure 4.5. The percentage of the relationship between smoking and eyesight

From Figure 4.5 above, it can be explained as follows:

- EL1R1 means smoking, there is an eyesight left response>1.0 and eyesight right response>1.0, which has a percentage of 48.5% which is the largest percentage.
- EL2R2 means smoking, there is an eyesight left response>1.0 and eyesight right response<1.0, has a percentage of 6.6%
- EL3R3 means smoking, there is an eyesight left response<1.0 and eyesight right response>1.0, which has a percentage of 6.1% which is the smallest percentage.
- EL4R4 means smoking, there is an eyesight left response<1.0 and eyesight right response<1.0, has a percentage of 38.8%.

The relationship between smoking and triglyceride and LDL

Researchers split data and then present the relationship between smoking and triglyceride and LDL in the form of pie charts and provide labels such as AT1L1, AT2L2, AT3L3, and AT4L4. Visualization can be seen in Figure 4.6 below:

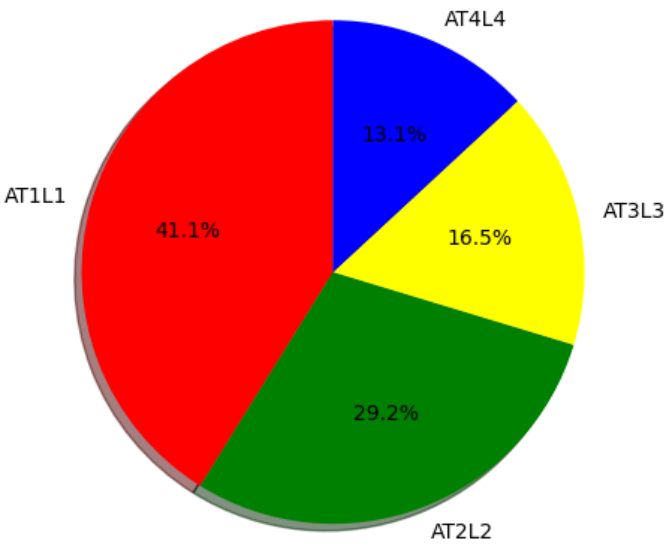

Figure 4.6. The percentage of the relationship between smoking and triglyceride and LDL

From Figure 4.6 above, it can be explained as follows:

- AT1L1 means smoking, there is a triglyceride response<150 and LDL response<100, which has a percentage of 41.1% which is the largest percentage.
- AT2L2 means smoking there is a triglyceride response<150 and LDL response>129, has a percentage of 29.2%.
- AT3L3 means smoking, there is a triglyceride response>199 and LDL response<100, which has a percentage of 16.5%.
- AT4L4 means smoking, there is a triglyceride response>199 and LDL response>129, has a percentage of 13.1% which is the smallest percentage.

4. CONCLUSION

The biggest percentage is found at a blood pressure of 44.4% on TS4R4 there is a Systolic Response <120 and relaxation <80, the blood sugar of 58.2% on FBS1 there is a Fasting Blood Sugar response <= 99, hemoglobin of 47.7% on MH3 means smoking, there is a response of hemoglobin> 17.2, oral and tartar by 62% on OYTY means smoking, there are oral responses and tartar responses, eyesight by 48.5% on EL1R1 means smoking, there is a response of eyesight left> 1.0 and eyesight right> 1.0, triglyceride and LDL by 41.1% on AT1L1 means smoking, there is a response of triglyceride> 150 and LDL< 100.

5. RECOMMENDATION

For now, the study only focuses on visualizing the relationship of smoking relationships with blood pressure, blood sugar, hemoglobin, oral, and tartar then may be visualized the relationship data using different body signals.

REFERENCES

[1] A. D. Mulyanto, "A Python-based Program for Performance Analysis of the Head of LP2M," JICTE (Journal Inf. Comput. Technol. EDUC., Vol. 6, No. 2, PP. 56-60, 2022, Doi: 10.21070/JICTE.V6I2.1618.

[2] Farrah Meirisah and Tata Sutabri, "Analysis of cybercrime cases using data visualization," J. Inform. Technology. and Science, Vol. 5, no. 1, PP. 32-37, 2023, Doi: 10.51401/Jinteks.v5i1.2211.

[3] D. Communication, S. K. Denpasar, K. Denpasar, K. Creative, B. Culture, and M. Denpasar, "Analysis and Visualization of Community Complaints Data from January to May on the application," Vol. 2, no. November, PP. 1-8, 2023.

[4] R. S. Oktavian and S. Budi, "Analysis of the Google Play Store Dataset uses the Exploratory Data Analysis method," J. Strateg. Maranatha, Vol. 2, no. 2, PP. 636–649, 2020.

[5] P. Lestari, I. Tahyudin, R. Waluyo, S. Information, F. I. Computer, and U. A. Purwokerto, "Analysis of Exploration and Visualization of Dividend for," Vol. 12, NO. 2, 2023.

[6] R. Adi Pranata et al., "Python Analysis The use of music as a treatment for mental disorders," Jupti, Vol. 2, no. 2, PP. 39–47, 2023.

[7] I. K. N. Adi Jaya, F. Fauzi, A. Suryana, A. D. Widiantoro, and I. D. K. L. Digita, "Data Visualization of House of Worship Distribution in the IKN Nusantara Region Using Python," J. Ilm. Merpati (Tower of Research. Agreement. Technology. Information), Vol. 11, NO. 1, p. 1, 2023, DOI: 10.24843/JIM.2023.V11.I01.P01.

[8] R. Maringka and K. Kusnawi, "Exploratory Data Analysis Factor of Mental Health Effects at Work," Cogito Smart J., Vol. 7, no. 2, PP. 215-226, 2021, Doi: 10.31154/Cogito.V7I2.312.215-226.

[9] S. Junaidi, M. Devegi, and H. Kurniawan, "Training on Processing and Visualization of Population Data Using Python," J. Servant. and empowerment. Masy., Vol. 4, no. 1, PP. 151–162, 2023, DOI: 10.30812/ADMA.V4I1.2963.

[10] D. Meliyana and K. Latifah, "Management and visualization of data at the Pati Regency Communication and Information Office," SCI. Eng. Natl. Semin., Vol. 7, no. 7, 2022.

[11] Kelly Hermanto, Darrius Salim, Bryan Wu, Odelia Regina Salim, and Ruby Belinda Gunadi, "The use of Python to analyze the pattern of the distribution of COVID-19 in Pandemi," J. Student Dev. Inf. System., Vol. 3, no. 2, PP. 62–75, 2023.

[12] R. Al Ghivary, M. Mawar, N. Wulandari, N. Srikandi, and A. N. M. F, "The Role of Data Visualization to Support Population Data Analysis in Indonesia," Pentehalix, Vol. 1, no. 1, p. 57, 2023, Doi: 10.24853/Penta.1.1.57-62.

[13] A. A. Prayogi et al., "Design and Visualization of Data from the Yogyakarta City DPRD Election Results," J. Teknol. Inf. and Comput., Vol. 9, no. 3, PP. 254-259, 2023.

[14] V. Lisia, A. E. Widjaja, A. R. Mitra, C. A. Haryani, and Hery, "Visualization of Geological Disaster Data in Web-Based Indonesia," Inf. System. Dev., Vol. 7, no. 1, PP. 9-27, 2022.

[15] M. B. Tamam, "Indonesia and Malaysia Data Visualization of the Spread of Covid 19 in Indonesia and Malaysia," Vol. 11, NO. 1, PP. 13-18, 2022.

[16] D. C. U. Lieharyani and R. Ambarwati, "Visualization of Tweet Data in the Higher Education Sector during the Pandemic Period," Build. Informatics, Technol. SCI., Vol. 4, no. 1, PP. 116–123, 2022, Doi: 10.47065/Bits.V4I1.1551.

[17] M. C. Kirana, N. P. Perkasa, M. Z. Lubis, and M. Fani, "Visualization of the Quality of Earthquake Information Dissemination in Indonesia Using Twitter," J. APPL. Informatics Comput., Vol. 3, no. 1, PP. 23-32, 2019, Doi: 10.30871/Jaic.V0i0.1246.
[18] E. Malang, B. Malang, and F. Malang, "A visualization of social media analytics on the social media account of the City of Kediri," Vol. 6, no. 2, PP. 183-192, 2022, [online]. Available: http://ejournal-binainsani.ac.id/index.php/imbi/article/view/1905%0ahttps://ejournal-binainsani.ac.id/index.php/imbi/article/download/1905/ 1455
[19] GATS Survey Findings: Adult smokers in Indonesia have risen the last 10 years. accessed on December 13, 2023 from https://sehatnegeriku.kemkes.go.id/baca/umum/20220601/4440021/temuan-survei-gats-perokok-dewasa-di-indonesia-naik-10-tahun-terakhir/
[20] 20/25 Vision: Good or Bad. accessed on December 13, 2023 from https://myvision-org.translate.goog/eyesight/20-25-vision/?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=tc
[21] Check fasting glucose. accessed on December 13, 2023 from https://www.halodoc.com/kesehatan/cek-glukosa-puasa
[22] Recognize hemoglobin, function, and normal levels in the body. accessed on December 13, 2023 from https://www.halodoc.com/artikel/mengenal-hemoglobin-fungsi-dan-kadar-normalnya-pada-tubuh
[23] Dental Plaque: Understand the causes and ways to get rid of it from the experts!. accessed on December 13, 2023 from https://www.tanyapepsodent.com/tips-kesehatan-gigi/plak-dan-kerusakan-gigi/plak-gigi-pahami-penyebab-dan-cara-menghilangkannya-dari-ahlinya.html
[24] Know the Normal Cholesterol Levels for Women & Men According to Age!. accessed on December 13, 2023 from https://www.siloamhospitals.com/informasi-siloam/artikel/kadar-kolesterol-normal
[25] Need to know, this is normal blood pressure according to age. accessed on December 13, 2023 from https://www.halodoc.com/artikel/perlu-tahu-ini-tekanan-darah-normal-sesuai-tingkatan-usia