Chapter 2

Statistical Mathematical Model for Smart Cities with Focus on Health Indicators

CLARISSA FERNANDA CORREIA LIMA LOUREIRO, ANGELA MARIA ALVES, LEONARDO BRUSCAGINI LIMA, AND GABRIEL GOMES DE OLIVEIRA

Understanding city data is extremely important to understand the reality and plan to improve the quality of life of the inhabitants. This document presents a part of the statistical mathematical model developed for the application of smart cities indicators metrics of the Smart and Sustainable Cities Maturity Model in Brazil. The focus of this work is on health indicators in the area of smart and sustainable cities in Brazil.

2.1 Introduction

There is worldwide concern about the migration of people to urban areas. The term "smart cities" or "smart cities" emerged from the concern to understand how to sustainably and efficiently accommodate people in urban areas. In Brazil, people who live in urban areas already make up almost 90% of the entire Brazilian population [1, 2, 3, 4].

The United Nations (UN) developed the 17 millennium goals and Sustainable Development Goals (SDGs) and, with them, several goals for improvements in cities and the world. These objectives are divided into themes, with their respective areas. To make it possible to measure the living conditions of cities and enable a diagnosis of the reality in which they find themselves in each of the themes, the information was divided, and, for better organization, some smart cities indicators were collected [5, 4, 6].

There are thousands of city indicators. However, the data from the indicators need to bring information that reflects properly and reliably the reality of the cities. In Brazil, smart city indicators are diverse, coming from different sources, with different metrics [7, 8, 9].

Understanding the current state of cities, with their problems, their capacity, and their possibilities for improving the quality of life of the people who inhabit them, as well as income generation equitably, it is necessary to understand the data. For this, it is necessary to translate the data to have effective information about the city [10, 11].

Mathematics and statistics make an intermediation that provides us with information translated from the data, a translation made by statistical concepts and methods. Statistics is the part of applied mathematics that provides methods for the collection, organization, description, analysis, and interpretation of quantitative data and their use for decision-making [12, 13, 14, 15].

It is important to understand how the data were collected, which variables were involved, if they are dependent on which areas, if they are quantitative or qualitative, and their types, to understand the reliability of the data and if they reflect the reality of the cities [16, 17].

In this article, the importance of a mathematical and statistical treatment of the data is explained, focusing on Brazilian cities for the development of a mathematical and statistical model that would be ideal for measuring all the dimensions of the Sustainable Development of Smart Cities Maturity Model in Brazil. Some data on health indicators will be highlighted, which are very important to understand and promote which public policies need to be developed and applied so that the quality of life of people in Brazil is better [18, 19].

2.2 Application of Mathematical Model for Public Health

Specifically in the health area, a mathematical model of indicators for smart cities can be useful to help managers understand and monitor the health of the population in real-time. This can help identify problem areas and develop interventions to improve public health.[20, 21, 22]

Many health indicators can be monitored in a smart city, such as chronic disease rates, air and water pollution levels, access to healthy food, and physical activity, among others. In Brazil, these indicators can be collected through official national bodies whose data are public, such as IBGE, INEP, and DATASUS, among other bodies with reliable and comprehensive data in the country [23, 24, 25].

A mathematical model can be developed to analyze this data and provide useful insights into the health of the population. For example, the model can identify areas with high rates of chronic disease and suggest specific interventions to improve health in those locations. The model can also be used to predict the spread of infectious diseases and identify areas that are most at risk of outbreaks [26, 27, 28, 29].

In addition, the mathematical model of health indicators can be used to assess the impact of specific interventions on population health. For example, if a smart city implements a physical activity promotion program, the model can be used to assess whether this intervention has had a positive impact on the population's health [30, 31].

In summary, a mathematical model of health indicators for smart cities can help improve public health and allow managers to make more informed decisions about health policies and specific interventions [32, 33, 34, 35, 36].

2.3 Methodology

Aiming at building a reliable metering that reflects the reality of the cities, research, studies, and the development of a statistical mathematical model were carried out for the metering part of the MMCISB (Portuguese acronym for Brazilian Sustainable Smart Cities Maturity Model) [32, 37].

To develop a good model, statistical analyses were made of the Brazilian secondary data, so that, individually, for each of the 80 indicators, the customized metric could be created and integrated into the entire model [32, 37].

First, the areas in which the statistical mathematical model would be applied were divided. The indicators were chosen with strict criteria that took into account the number of cities that had data, the maturity model of the International Telecommunications Union (ITU, 2016) the millennium goals, Sustainable Development Goals (SDGs) of the UN (UN, 2015), ISO standards (ISO, 2018), and Brazilian Smart Cities Charter (Charter, 2020), among other criteria [32, 38].

Of the four dimensions of the MMCISB, three dimensions were chosen because they are areas in which the indicators come from secondary data, that is, open data from cities in Brazil. The three dimensions of sustainable development are: economic, socio-cultural, and environmental. These dimensions have their topics and indicators [32, 38].

2.3.1 Indicator Analysis

A careful and detailed analysis of 80 secondary data indicators was carried out. Each indicator analysis is unique and specific to that indicator. The calculations, the graphs, and the reasoning were described in detail and documented in an extensive report, which is restricted to the Poli. Tic laboratory team, a CTI Renato Archer laboratory, specialized in promoting public policies and also works with information and communication technologies, to improve the quality of life of people living in Brazil [39, 40].

For each indicator, a detailed and careful analysis of the data was carried out with the following items:

- Statistical analysis;
- Normalization;
- Scale Transformation;
- Criteria for defining the level intervals;
- Statistical Weight;
- Indicator contribution formula for the maturity model;
- Model application report and examples for Developers;
- Outliers rules and equations.

For some indicators, technical notes were prepared to explain some references and information. Information about the indicators, the relationship between grade and quality of life, and the importance and relevance of each indicator are explained in detail in the methodology of the Maturity Model of Smart Cities in Brazil [39].

In the document mentioned above, all the metrics for each indicator, the contribution of the score for each topic, and the dimension were exposed. The integration of the four dimensions was done and well-detailed in the methodology of the MMCISB model. The order of each dimension, with the respective topics, was followed to proceed with the measurement of indicators for the entire area of Sustainable Development [39].

2.3.2 Outliers

In this work, we will not detail all the points of the analysis of the health indicators data. We will show, within statistics, the importance of analyzing indicators using artificial intelligence techniques. We will also explain the entire context of outliers, which are points outside the curve, exceptions, some data that are out of reality due to data imputation errors, some accidents, and system errors, among other types of errors [41, 42].

Outliers are data that differ from the standardization of a data set. These are values that deviate from normality, they are points outside the curve, out of the ordinary, and can cause anomalies in the results obtained from the analysis of the indicators. Understanding them is fundamental in data analysis because they can change the entire result of an analysis [41, 42].

As outliers can undermine any analysis of off-curve data, recommendations and observations can be made to cities about their outliers. Data can be excluded, for example, when the discrepant data is the result of an entry or typing error, in this case, the removal of the outlier from the sample is indicated [43, 44].

There are many methods for identifying outliers in academia, for example, the Tukey method or the boxplot that defines lower and upper limits, based on the interquartile (IQR) and the first and third quartiles. Data that fall outside these limits will be determined as an outlier. The categories of outlier detection methods can be, for example, methods based on distance, statistics, depth, density, and also deviation. The difference between each category lies in the way the method uses to separate common data from data considered outliers [43, 45, 44].

2.3.3 Standard Deviation Method for Outlier Detection

The standard deviation (σ or SD) method or free standard deviation is one of the most used, both in academia and in companies that deal with the area of data analysis. This method was chosen to detect outliers in the model and is depicted in Figure 2.1, which demonstrates that 68.3% of the values will be within the range $\mu \pm 1\sigma$, 95.4% will be within $\mu \pm 2\sigma$, and 99.7% will be within $\mu \pm 3\sigma$ [46, 47, 48].

The Standard Deviation method or free standard deviation chosen is based on the following formula:

$$X \pm 3(SD) \tag{2.1}$$

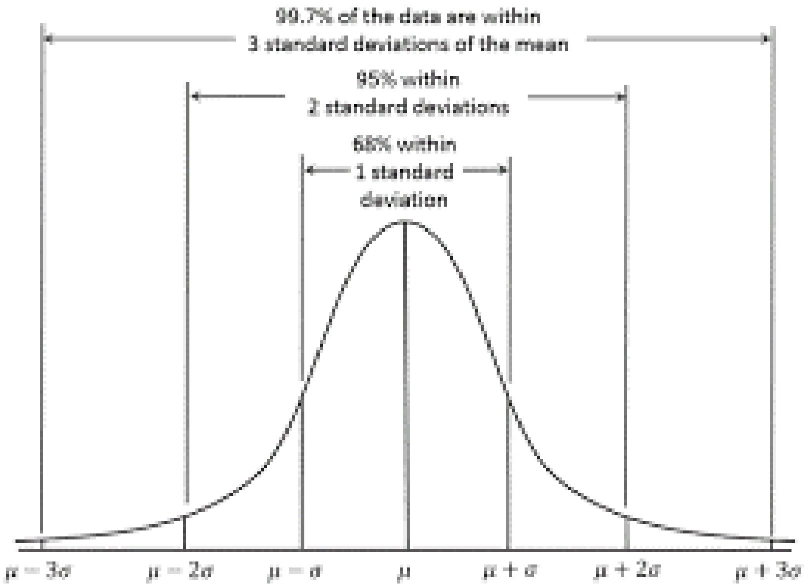

Figure 2.1: Standard deviation in a normal distribution.

where X is the average of the results of each indicator; SD is the standard deviation of the respective indicator.

Outliers are identified if their respective values are less than the smallest value in the formula when using the minus sign:

$$X - 3(SD) \tag{2.2}$$

Outliers are identified if their respective values are greater than the highest value when using the plus sign, defined by the formula:

$$X + 3(SD) \tag{2.3}$$

2.3.4 Steps for Calculating Outliers

To calculate the outliers, we can do it step by step with the data, as in an algorithm, or sequence:

- Simple Arithmetic Average;
- Variance Identification:

$$\frac{(x_1 - x)^2}{n-1} + \frac{(x_2 - x)^2}{n-1} + \frac{(x_3 - x)^2}{n-1} + \cdots + \frac{(x_n - x)^2}{n-1} \tag{2.4}$$

- Standard Deviation Identification: $\sqrt{var}$;
- Application of the formula: $X \pm 3(SD)$.

2.3.5 Indicator Criteria for Selecting the Calculation of Outliers

For each indicator, a statistical analysis was performed in which the averages, medians, first quartile, third quartile, standard deviation, and variance were calculated, for the normalization of the values of the indicators related to each maturity level of the indicators, applying the boxplot method [49, 50, 51].

In addition to the previous analyses, the Standard Deviation (SD) or Free Standard Deviation method was used to increase the reliability of the data, to inform the cities about results that are far outside the normality of the averages of the Brazilian data, recommending the cities to take appropriate action when necessary [49, 50, 51].

The indicators that will go through the Standard Deviation (SD) or Free Standard Deviation method, within the data analysis of the model, will be indicators with numerical data results, and that meet the following criteria [49, 50, 51]:

- Indicators cannot be binary, as they have only 2 response options;
- The indicators cannot have answers with the sum of alternative or multiple-choice scores. For example, indicators from the MUNIC database;
- The indicators cannot be from a primary source, as this is information that has not yet been acquired in the system, most of the indicators are the sum of the alternatives' scores, multiple choice, or binary, so they do not apply to the calculation, according to the above criteria;

2.4 Conclusion

In this work, the importance of a mathematical and statistical treatment of data, focusing on health indicators, of Brazilian cities was explained for the development of an ideal mathematical and statistical model to measure all the dimensions of the Sustainable Development of the Maturity Model of Smart Cities do Brasil, MMCISB, (CTI/poli.TIC, 2020)

All calculations, analyses, and treatment of data from the mathematical statistical model were put into practice in all indicators of the dimensions of Sustainable Development of the MMCISB. The model worked, all the formulas and analyses of each of the indicators were applied, and the platform developers could understand how to apply them to the algorithms for future work, which is the implementation of a platform for use in all Brazilian cities.

Bibliography

[1] Ricardo Alexandre Afonso, Kellyton dos Santos Brito, Clóvis Holanda do Nascimento, Vinicius Cardoso Garcia, and Alexandre Álvaro. Brazilian smart cities: using a maturity model to measure and compare inequality in cities. In *Proceedings of the 16th annual international conference on digital government research*, pages 230–238, 2015.

[2] Gabriel Gomes de Oliveira, Lucas Alves Rodrigues de Sá, Yuzo Iano, and Gabriel Caumo Vaz. Security in smart home using blockchain. pages 306–313, 2023.

[3] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Pablo David Minango Negrete, and Daniel Rodrigues Ferraz Izario. Horizontal curves with transition. the use of this methodology for the calculation of a road project in the city of campinas/sp-brazil. pages 51–65, 2022.

[4] Juliana P da S. Ulian, Luiz Carlos Pereira da Silva, Gabriel Gomes de Oliveira, João Guilherme Ito Cypriano, Yuzo Iano, and Gabriel Caumo Vaz. Telemanagement and its benefits to energy, environment, and society: A case study in street lighting. pages 178–187, 2022.

[5] HM Van Bellen. Sustainability indicators: A comparative analysis, fgv, rio de janeiro, 2005.

[6] Lilian Regis Laraia, Yuzo Iano, Ricardo Takahira, Luiz Vicente Figueira de Mello Filho, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Technology for electric bus in the brazilian scenario: Focus on the adoption of national components. pages 276–285, 2022.

[7] Johannes Mueller, Hangxin Lu, Artem Chirkin, Bernhard Klein, and Gerhard Schmitt. Citizen design science: A strategy for crowd-creative urban design. *Cities*, 72:181–188, 2018.

[8] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Pablo David Minango Negrete, and Juan Carlos Minango Negrete. Prop walls: A contextualization of the theme in a case study in the city of campinas (brazil). pages 128–139, 2022.

[9] Polyane Alves Santos, Yuzo Iano, Kelem Christine Pereira Jordão, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, Ingrid Araújo Sampaio, and Euclides Lourenço Chuma. Analysis of the relationship between maturity indicators using the multivariate linear regression: A case study in the brazilian cities. pages 203–210, 2022.

[10] Janaina Pamplona da Costa et al. *Network (Mis) Alignment, Technology Policy and Innovation: The Tale of Two Brazilian Cities*. SSRN, 2016.

[11] Ingrid Araújo Sampaio, Yuzo Iano, Aurelio Ribeiro Leite de Oliveira, Lino Marcos da Silva, Rinaldo Vieira da Silva Júnior, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Polyane Alves Santos, and Kelem Christine Pereira Jordão. The use of the elman preconditioner in the early iterations of interior point methods. pages 355–363, 2022.

[12] Paul M Thompson, Jason L Stein, Sarah E Medland, Derrek P Hibar, Alejandro Arias Vasquez, Miguel E Renteria, Roberto Toro, Neda Jahanshad, Gunter Schumann, Barbara Franke, et al. The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior*, 8:153–182, 2014.

[13] Everton Hideo Nishimura, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Application and requirements of aiot-enabled industrial control units. pages 724–733, 2022.

[14] Gabriel Caumo Vaz, Yuzo Iano, and Gabriel Gomes de Oliveira. Iot-from industries to houses: An overview. pages 734–741, 2022.

[15] Celso Fabricio Correia de Souza, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Valéria Sueli Reis, and Josué Mastrodi Neto. Institutional development index (idi): Calculation for municipalities in the metropolitan region of campinas (brazil). pages 245–255, 2022.

[16] Seth Schindler and Simon Marvin. Constructing a universal logic of urban control? international standards for city data, management, and interoperability. *City*, 22(2):298–307, 2018.

[17] Domingos Teixeira da Silva Neto, Jéssica Fernandes Alves, Polyane Alves Santos, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Yuzo Iano, and Lucas dos Santos Ribeiro. Proposal mppt algorithm using the kalman filter. pages 750–759, 2022.

[18] Leonardo Bruscagini de Lima, Yuzo Iano, Pedro Y Noritomi, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Data security, privacy, and regulatory issues: A conceptual approach to digital transformation to smart cities. pages 256–263, 2022.

[19] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Pablo David Minango Negrete, and Juan Carlos Minango Negrete. Structural analysis of bridges and viaducts using the iot concept. an approach on dom pedro highway (campinas-brazil). pages 108–119, 2022.

[20] Negar Noori, Martin de Jong, Marijn Janssen, Daan Schraven, and Thomas Hoppe. Input-output modeling for smart city development. *Journal of Urban Technology*, 28(1-2):71–92, 2021.

[21] Leonardo Bruscagini de Lima, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Alecssander Daniel de Almeida, Gustavo Bertozzi Motta, Gabriel Matsumoto Villaça, Matias Oliveira Schwarz, and Pedro Y Noritomi. Mathematical modeling: A conceptual approach of linear algebra as a tool for technological applications. pages 239–248, 2022.

[22] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Pablo David Minango Negrete, Juan Carlos Minango Negrete, and Euclides Lourenço Chuma. Intelligent mobility: A proposal for modeling traffic lights using fuzzy logic and iot for smart cities. pages 302–311, 2022.

[23] Nutthawut Ritmak, Wanchai Rattanawong, and Varin Vongmanee. A new dimension of health sustainability model after pandemic crisis using structural equation model. *Sustainability*, 15(2):1616, 2023.

[24] Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Lourenço Chuma, Roger Prior Gregio, and Alessandra Cristina Santos Akkari. Analysis of the ergonomic concept of public transportation in the city of campinas (brazil). pages 453–459, 2021.

[25] Telmo Cardoso Lustosa, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, and Valéria Sueli Reis. Safety management applied to smart cities design. pages 498–510, 2021.

[26] Xiangjie Kong, Menglin Li, Tao Tang, Kaiqi Tian, Luis Moreira-Matias, and Feng Xia. Shared subway shuttle bus route planning based on transport data analytics. *IEEE Transactions on Automation Science and Engineering*, 15(4):1507–1520, 2018.

[27] Roger Prior Gregio, Yuzo Iano, Lia Toledo Moreira Mota, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, Diego Arturo Pajuelo Castro, and Carolina Fernandes Frangeto. Energy use in urban areas using neodymium magnets. pages 988–1005, 2021.

[28] Euclides Lourenco Chuma, Yuzo Iano, Leonardo Lorenzo Bravo Roger, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Novelty sensor for detection of wear particles in oil using integrated microwave metamaterial resonators with neodymium magnets. *IEEE Sensors Journal*, 22(11):10508–10514, 2022.

[29] Y Thiagarajan, Baburao Pasupulati, Gabriel Gomes de Oliveira, Yuzo Iano, and Gabriel Caumo Vaz. A simple approach for short-term hydrothermal self scheduling for generation companies in restructured power system. pages 396–414, 2022.

[30] Daniel Katz Bonello, Yuzo Iano, Umberto Bonello Neto, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. A study about automated optical inspection: Inspection algorithms applied in flexible manufacturing printed circuit board cells using the mahalanobis distance method 1. pages 198–212, 2022.

[31] Raed Fawzi Mohammed Ameen and Monjur Mourshed. Urban sustainability assessment framework development: The ranking and weighting of sustainability indicators using analytic hierarchy process. *Sustainable Cities and Society*, 44:356–366, 2019.

[32] Jinsong Wu, Song Guo, Huawei Huang, William Liu, and Yong Xiang. Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives. *IEEE Communications Surveys & Tutorials*, 20(3):2389–2406, 2018.

[33] Daniel Izario, João Brancalhone, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, and Karine Izario. 5g-automation of vertical systems in the industry 4.0. pages 35–43, 2022.

[34] Antonio Carlos Demanboro, David Bianchini, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. 6g networks: An innovative approach, but with many challenges and paradigms, in the development of platforms and services in the near future. pages 172–187, 2022.

[35] Y Thiagarajan, G Palanivel, ID Soubache, Gabriel Gomes de Oliveira, Yuzo Iano, Gabriel Caumo Vaz, and Himanshu Monga. Design and fabrication of human-powered vehicle-a measure for healthy living. pages 1–15, 2022.

[36] Alex Restani Siegle, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Proposal of mathematical models for a continuous flow electric heater. pages 213–228, 2022.

[37] Paolo Rodrigo de Oliveira Bacega, Yuzo Iano, Bruno Campos Simoni de Carvalho, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, and Euclides Lourenço Chuma. Study about the applicability of low latency in has transmission systems. pages 73–87, 2022.

[38] Pablo Minango, Yuzo Iano, Euclides Lourenço Chuma, Gabriel Caumo Vaz, Gabriel Gomes de Oliveira, and Juan Minango. Revision of the 5g concept rollout and its application in smart cities: A study case in south america. pages 229–238, 2022.

[39] Rui Chen, Wen-Xuan Long, Guoqiang Mao, and Changle Li. Development trends of mobile communication systems for railways. *IEEE Communications Surveys & Tutorials*, 20(4):3131–3141, 2018.

[40] Y Thiagarajan, Gabriel Gomes de Oliveira, Yuzo Iano, and Gabriel Caumo Vaz. Identification and analysis of bacterial species present in cow dung fed microbial fuel cell. pages 16–24, 2022.

[41] UNECFE UNECE and ITU ITU. Key performance indicators for smart sustainable cities to assess the achievement of sustainable development goals. *Draft New Recommendation ITUT L. 1603 (ex L. KPIs-SSC-SDGs)*, 2016.

[42] Juan Carlos Minango Negrete, Yuzo Iano, Pablo David Minango Negrete, Gabriel Caumo Vaz, and Gabriel Gomes de Oliveira. Sentiment and emotions analysis of tweets during the second round of 2021 ecuadorian presidential election. pages 257–268, 2022.

[43] Edward B Barbier and Joanne C Burgess. Sustainable development goal indicators: Analyzing trade-offs and complementarities. *World development*, 122:295–305, 2019.

[44] Adolfo Blengini Neto, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Fabiana Silva Podeleski, Higor de Paula Kolecha, and Marcius FH de Carvalho. The bfs method in a cloud environment for analyzing distributed energy resource management systems. pages 349–362, 2022.

[45] Paulo de Martino Jannuzzi. Indicadores para diagnóstico, monitoramento e avaliação de programas sociais no brasil. 2005.

[46] Bernard Rosner. *Fundamentals of biostatistics*. Cengage learning, 2015.

[47] Leonardo Bruscagini de Lima, Yuzo Iano, Gabriel Gomes de Oliveira, Gabriel Caumo Vaz, Alecssander Daniel de Almeida, Gustavo Bertozzi Motta, Gabriel Matsumoto Villaça, Matias Oliveira Schwarz, and Pedro Y Noritomi. Mathematical modeling: A conceptual approach of linear algebra as a tool for technological applications. pages 239–248, 2022.

[48] Juan Carlos Minango Negrete, Yuzo Iano, Pablo David Minango Negrete, Gabriel Caumo Vaz, and Gabriel Gomes de Oliveira. Sentiment analysis in the ecuadorian presidential election. pages 25–34, 2022.

[49] Giego Alves de SOUZA et al. Sistema de medição e indicadores de desempenho no ifmg: estudo descritivo e proposta de intervenção. 2019.

[50] Antonio Carlos Demanboro, David Bianchini, Yuzo Iano, Gabriel Gomes de Oliveira, and Gabriel Caumo Vaz. Regulatory aspects of 5g and perspectives in the scope of scientific and technological policy. pages 163–171, 2022.

[51] Gabriel Gomes De Oliveira, Yuzo Iano, Gabriel Caumo Vaz, Euclides Loureno Chuma, and Rangel Arthur. Intelligent transportation: Application of deep learning techniques in the search for a sustainable environment. pages 7–12, 2022.