creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

석사 학위논문

Artificial Intelligence applied to smart cities selection for e-waste processing units installation

전자 폐기물 처리 장치 설치를 위한 스마트 시티 선정에 인공 지능 적용

2021년 8월

서울대학교 대학원

공과대학

International Technology Professional Program

Marcos Pereira de Novais

Artificial Intelligence applied to smart cities selection for e-waste processing units installation

지도교수 Hwang Junseok

이 논문을 공학 석사학위 논문으로 제출함

2021 년 8 월

서울대학교 대학원

협동과정 기술경영경제정책 전공

Marcos Pereira de Novais

마르코스의 석사학위 논문을 인준함

2021 년 8 월

위 원 장 Jörn Altmann

부위원장 황준석

위 원 윤현영

Abstract

Artificial Intelligence applied to smart cities selection for e-waste processing units installation

Marcos Pereira de Novais

International Technology Professional Program

School of Engineering

Seoul National University

E-waste generation has been a growing problem globally, as e-waste management and its recycling unit capacity are not capable of giving proper disposal to that volume. The Smart Cities Index (SCI) can contribute to e-waste disposal and the resolution of the problems by using economic, social, sustainable, technological, and legislation data. The research proposes a data analysis of the relation between the Smart City features and the e-waste related indicators by using the approach of Artificial Intelligence unsupervised algorithm of K-means clustering, and the Generalized Linear Model of prediction, comparing the models and extracting the Smart City key features, to the determine the installation of a new e-waste recycling processing unit, the results and its implication, where it was observed the high correlation between the e-waste management and the Smart City characteristics.

Keywords: E-waste management, Smart City, Artificial Intelligence, Clustering, Prediction, Smart City Index, Sustainability

Student Number: 2019-23118

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Smart cities and E-waste management

Smart cities are considered potential solutions for global impacts such as global warming, resource scarcity, health, and population growth. The “smartness” can be defined from a set of indicators, e.g., technological, interconnected, sustainable, comfortable, safe (Grossi & Trunova, 2021). Although, these authors emphasize the complexity and ambiguity of smart city definition, the importance of Information, Communication and Technology (ICT) tools applied to sustainability is highlighted as prevalent in the literature (Benites & Simões, 2021; Hoang, Pham, & Nguyen, 2021; Shamsuzzoha, Nieminen, Piya, & Rutledge, 2021).

Solid waste management (SWM) is one of the main challenges of smart cities (Benites & Simões, 2021; Cheela, Ranjan, Goel, John, & Dubey, 2021) as a result of urbanization, resources demand and products consumption. Sharma et al. (2020) highlight the importance of proposing a sustainable ecosystem to minimize the impacts of urbanization, such as providing better infrastructure and transportation for waste management. Esmaeilian et al. (2018) highlight the importance of waste management inside the environmental and sustainable focus of smart cities and propose tracking and data sharing technologies to reduce waste generation and enhance waste recovery. Other authors took the same road and presented technological solutions for waste management in different countries.

The circular economy concept emerges as an interdisciplinary approach that aims to promote the recovery of value from restorative and regenerative systems. Thus, the encouragement of reuse and the recovery of materials and products from post-consumption, are prioritized. According to Arora, Paterok, Banerjee, and Saluja (2017), the circular economy has, among its tools, urban mining, consisting of resources recovering from anthropogenic mines, such as landfill, sewage, ash from incineration and e-waste. Thus, the transition from a model based on primary mining moves towards incorporating urban mining, also known as secondary mining.

The technological innovations improvement has allowed important benefits but also created a problem of global scale, the generation of waste electrical and electronic equipment (WEEE or e-waste), that rises 3 to 5% each day, more than other usual waste (Forti, Baldé, Kuehr, & Bel, 2020).

According to the Global E-waste Monitor, in 2019 53.6 million metric tons of e-waste were generated, with a 7.3kg average per capita, an increase of 19% from 2016. The global generation of e-waste in 2014 was 44.4 million metric tons, and the expected volume is 74.7 million metric tons of e-waste to be generated in 2030; this represents almost double the of e-waste generated in 16 years (Forti et al., 2020).

Due to the potential for recovering valuable materials, e-waste management represents an important source of secondary resources such as copper, gold, platinum, and silver. Forti et al. (2020) estimate a recovery potential of about 57 billion dollars to recover valuable materials from e-waste. However, the recycling market has a high degree of uncertainty, especially related to the supply chain and the availability of secondary materials.

One of the challenges faced by e-waste management is the location of material recycling facilities (Cheela et al., 2021; Dutta & Goel, 2021; Song et al., 2019). The location of e-waste processing units can result in human and environmental health impacts, as it can solve logistics and optimize routes. Good waste management practices must always be prioritized to guarantee the balance between sustainability and profitability.

The dispersed locations of recycling centers are also a problem that makes the recycling market doubt its potential value and capacity. Most developing countries and cities with big areas of land need to understand and invest in more recycling units, especially for e-waste, considering its growth and toxicity (Abbondanza & Souza, 2019).

To Veiga (2013), the increased industry cost of using recycling material put the recycling market down. It needs to be improved to acquire competitiveness and increase its value as a product. To do so , it is needed to improve and expand the recycling industry.

Achieving that would not be easy, but once it is reached, it would bring innovation, working almost disassociated from the legislation, with a growth rate of capability increasing without effort, like the aluminum market, where "the value of the material recycled is economically viable" (Veiga, 2013).

Usually, the households are located in larger cities, which can make the price of e-waste material higher, decreasing and impacting its market. The transportation and the lack of a proper infrastructure near collection points also drive the value of the e-waste to decrease its competitiveness (Shevchenko et al., 2021).

An efficient and proper installation of a recycling center brings innovation to its city, decreasing the e-waste recycling value and bringing more competitiveness to its local market. To decrease the value of the e-waste to be processed, the industry must be efficient, with a high rate of recycling and lower impact on the environment, for that, the city must have a proper infrastructure to receive the recycling center, offering characteristics to support and improve the innovation through efficiency, as a smart city can provide.

Considering the smart management approach of smart cities and the need for an advanced location to improve the value of the e-waste market, the analysis of the installation of an e-waste recycling unit based on Smart City characteristics can help to archive that goal faster, increasing the rate of recycling products, with lower environmental impact, and improving the e-waste recycling market with driven innovation.

Thus, in the next topics, the aspects that underlie e-waste management and aspects of the consolidation of smart cities will be presented as elements for discussion.

1.2 Background

According to Arora et al. (2017), different materials can be recovered from waste landfills and dumps, building stock, end-of-life vehicles, municipal solid waste (MSW) and electronic waste (E-waste) through urban mining techniques. While landfills and dumps streams are more fitted to waste-to-energy approaches, end-of-life vehicles (ELV) and e-waste are important sources of valuable materials.

The Global E-waste Monitor also reveals that only 17% of all the 2019 e-waste was properly treated; from all the e-waste, it could be retrieved up to $ 57 billion USD, considering that all e-waste contains valuable material, such as copper, iron, gold, and even aluminum (Forti et al., 2020).

This 17%, around 9.3 metric tons, represents around $10 billion recovered in raw material, generating 4 million metric tons. Another benefit is the energy saving from not producing or extracting the same amount of raw material, decreasing the CO2 emission by 15 million tons, only by using the secondary raw materials of copper, aluminum, and iron, recovered in 2019 (Forti et al., 2020).

The e-waste lifespan is also decreasing, not because it stops working, but because of the rapid innovation and consumer tendency to update the technology for a brand-new one, especially in developed countries.

E-waste contains hazardous substances that can cause pollution to the environment; if the e-waste disposal and treatment are not done properly, it can produce a huge impact on the land, decreasing the value of the recycling industry and market.

To avoid this, the level of innovation and readiness of the city must be advanced, placing the recycling market into another level of maturity, acquiring all its potential by the right selection of the location of the e-waste recycling unit.

E-waste must be handled and treated properly, where its management will help improve the local market and prevent the contamination of the landfill and citizens. Many procedures and methodologies of treatments exist and can be applied to handle this kind of e-waste.

For many years, the Life Cycle Assessment (LCA) has been studied and used by countries to define and manage e-waste, but this does not cover the Circular Economy aspect of it. The main goal of LCA is to describe in detail the production life cycle of a product, measuring the impact of the process on the environment, register all the material and services used and the emissions caused by it.

E-waste management is not related to the product as a unit like LCA, but it covers all the processes of managing the generated e-waste by implementing legislation and responsibilities to ensure that it will have a proper final disposal. Unfortunately, e-waste management is still deficient in most countries, especially developing countries (de Oliveira, Bernardes, & Gerbase, 2012).

The differences are based on proper infrastructure, legislation, and technology; while the developed countries are already working with extended producer responsibility (EPR) and/or with the reverse logistic system, most developing countries did not establish the e-waste legislation that covers those responsibilities in their full aspect.

It is important to consider that developed countries have more capable e-waste recycling processing units installed in their cities, especially in smart cities with efficient waste management, and can invest much more in Research and Development (R&D).

1.2.1 Global panorama

Countries like Japan and Switzerland implemented the LCA combined to proper regulations to develop an optimal model to manage the e-waste (Doan, Amer, Lee, & Phuc, 2019).

China also developed, since 2009, regulations to improve its e-waste management, promoting enterprises by giving subsidies to the public on purchases of new products by the policy “Old for New.” This regulation enhanced the e-waste management of China, increasing the recycling rate up to 84.5% of all e-waste produced in 2011 (Cao et al., 2016), which was possible because of the presence of advanced e-waste recycling centers, proper city infrastructure, and the proper legislation.

The Bloomberg Innovation Index 2021 index (Figure 1) shows South Korea in the 1st place among the Most Innovative Economies of the world; one of the main reasons is the Smart City investments and innovation throw a greener environment, highlighting the importance of the relationship between environmental sustainability and the establishment of criteria that define sustainable cities.

Figure 1. Bloomberg Innovation Index 2021 map (Bloomberg 2021)

This greener environment is taking place from actions like the Korea New Deal, which contains the National Strategy for a Great Transformation, including projects such as the Green New Deal, with the Circular Economy plan, that intends to solve the problem of the recycling rate of 12%, with the perspective to increase it, and apply the innovation and smart infrastructure, promoting the interconnection of waste recycling between companies, stimulating the use of raw material and energy by other companies (Lee & Woo, 2020).

This project focuses on the green transition of infrastructure, low-carb and decentralized energy and Innovation in the Green Industry, putting South Korea in the leading for High-end digital & IT technology that can be integrated into the green industry, including e-waste management (Korea, 2020).

Brazil is the biggest country in South America, with and most promising emerging economy that has already taken some action in the direction of the e-waste management, with regulations for the Extend Producer Responsibility (EPR) legislation released in the form of a Presidential Decree, published in February of 2020, that regulates the reverse logistics of electronics products, following the sectoral agreement with entities representing the main electronics companies in the country as a way to enforce the reverse logistics (Brasil, 2020). This Presidential Decree is based on the law that instituted the Brazilian Solid Waste Policy (Law No. 12,305/2010 and Decree No 7,404/2010), created to pursue the sustainability and circular economy of the country, but this alone did not promote the e-waste management in Brazil; one of the motives of the Presidential Decree, the increase of certified recycling unit is needed, but with the budget limited, the selection of the right city will help to faster increase the recycling rate (Xavier, Ottoni, & Lepawsky, 2021).

Like many cities in Brazil, Rio de Janeiro is one of the most advanced cities in the direction of proper e-waste management, producing approximately 127 tons of e-waste every day. It has 251 collections points and 24 e-waste recycling industries capable of proper recycling (Ottoni, Dias, & Xavier, 2020).

Rio de Janeiro and São Paulo are the two cities in Brazil that are listed in the Smart City Index, even having deficiency on information about the individual production, proper treatment, disposal, reuse, and collection of the e-waste, they are model, considering the e-waste recycling rates, and when selecting a location to install a recycling unit.

The Smart City Index is a "*Holistic attempt to capture the various dimensions of how citizens could consider that their respective cities are becoming better cities by becoming smarter ones"* (Center, 2020).

The report is also the result of surveys applied to the citizens of the 109 cities mentioned on it, where they expressed their perception of the smartness of their cities.

This report contains aspects such as education, safety, health, economic, life expectation, environmental respect, mobility, and technology. E-waste management can be found inside many of those aspects and creates the perception of a strong relationship between the Smart City and successful e-waste management.

This relationship can be more understandable when considering that circular economy, as a maximized efficient use of resources and waste, is present in both concepts, not only being a core of the e-waste management, but also as part of the philosophy of Smart City implementation.

There is no Smart City without maximized efficiency, and e-waste management helps to achieve that. Therefore, there will not be improvements without the implementation of TIC technology, such as the e-waste processing units.

Having an E-waste processing unit installed on the city can improve the green goals by reducing the CO2 emissions and saving energy, as much as improving the innovation. One example is the study case of Ukraine of the implementation of smart e-waste reverse system for cost minimization, the study shows "*a solid rationale for involving local e-waste operators as key stakeholders of the smart e-waste reverse system*" (Shevchenko et al., 2021).

1.3 Purpose of Research

Legal requirements mostly drive e-waste management. The main aspects of e-waste hazardousness, impacts, amount generation, and collection rates were established from the initial international documents, e.g., Basel Convention (Choksi, 2001) and the first version of the European Directives.

Nevertheless, despite the European Directives, these aspects are not harmonized from country to country, and in some cases, as in the United States, even from city to city.

To mitigate the impact resulting from the intensive growth of e-waste generation based on a metric common to countries, it is proposed to evaluate smart cities in terms of their potential to contribute to efficient and sustainable management of e-waste, and therefore helping to predict a better place to implement the recycling units, improving the recycling market and promoting the decrease of the e-waste as a product of purchase.

Both collection and processing are the bottlenecks of e-waste treatment (Ottoni et al., 2020); while collection requires an efficient reverse logistics system, the treatment requires specific processes. However, either collection and treatment have a feedback relation, and as long as material is inputted in the system, the more efficient and cost-effective the entire system is. The efficiency reinforces the sustainability parameters and endorses the material input, as well.

To acquire a better e-waste treatment, the installation of an e-waste processing unit, in a smart city context, can contribute to the circular economy of cities, providing resources to be analyzed.

Therefore, to understand the relation between the smart city index and the e-waste management, this research has the proposal of clarifying "*What are the key indicators of smart cities that makes them satisfactory location of e-waste management facilities?*" and if "*Is there any similarity between the classification and prediction models that helps to identify the correlations between the e-waste management and the Smart City Index?*"

Chapter 2 Literature review

As the key factor, the technologies of smart cities have an important role in decreasing greenhouse gas emissions and improving the energy efficiency of the city (Ahvenniemi, Huovila, Pinto-Seppä, & Airaksinen, 2017).

These characteristics are embedded in innovation and improvement of the way the citizens perceive life and the surroundings. In that direction, to select which smart city factor can contribute to determining the optimal location of an e-waste recycling unit, it is important to understand and know the main factors.

With this in mind, to build a literature review to answer the research questions, some key expressions and keywords were used. The first searches were based on the terms "e-waste management" and "smart city index" in all available fields and for the period of the last 10 years. From the SNU Library website[1], the search resulted in 126 articles that were refined and selected by their relevance to the research topic.

Another search for the related topic was the use of terms "smart city," "key

[1]The SNU library website (https://library.snu.ac.kr/) contains access to more than 300 databases, including Web of Science, OCDE, Scopus, and many others.

indicators," and "e-waste recycling center," to understand if the question was answered in previous literature. The result from the SNU Library website considering the last 10 years, showed 68 articles that were refined and selected by their relevance to the research, showing that the topic was not yet discussed. At last, the search on the SNU library website was made with the terms: Smart city; E-waste recycling unit; Key indicators; Artificial Intelligence; Prediction; and Clustering; the results showed only one article, "*Methodological Proposals for the Development of Services in a Smart City: A Literature Review*" by Serey et al. (2020).

Serey et al. (2020) divided the Smart Cities trend into Domains, such as Government, Environment, Urban Settlement, Social Service, and Economy, describing and listing the AI applications areas on the smart city technologies, giving keys indicators for all domains, where we can highlight Waste management, Environmental road infrastructure, Environmental control, City monitoring and City management as items with key indicators.

The Serey et al. (2020) literature review reinforces the Smart City's concept of innovation and efficiency by demonstrating the technologies that increase the capability of administration and manageability, not only that but also demonstrate that "*The circular economy of interconnected areas allows positive business results to be intensified*," for example.

Once again, the literature shows some aspects to be considered in the study, the economic situation, the circular economy, the e-waste management regulations, system, framework, and technology, the Smart City Index, and its parameters, as well as the presence of resources to receive and transport the e-waste in a more efficient way.

The fact that Smart Cities uses innovative solutions to manage the waste makes it be the perfect standard to determine the location of one more efficient e-waste recycling unit; for example, by using Internet of Things (IoT) sensors, one smart city can have smart collection services, that optimize the process, allowing it to create more efficient and economic collections routes (Gutierrez, Jensen, Henius, & Riaz, 2015), highlighting the economic and technological indicators as indispensable key factors.

By defining the scope of the search, the findings on previous studies made clear that the topic of creating a prediction model, based on specifics criteria of smart city and e-waste management relation led to broader aspects.

Those aspects were organized in indicators to facilitate the understanding the behavior of the last ten years researchers approaching the themes of e-waste management, smart city and prediction using secondary data.

Ultimately, the key factors of a Smart City towards e-waste management are inside its own concept, where the role of the ICT use on it demonstrates the goal of acquiring prosperity, effectiveness and competitiveness, while reaching an innovative, better and greener environment, with a higher quality of life (Ahvenniemi et al., 2017).

2.1 Social and economic indicators

Most of the literature about the social and economic indicators has a strong relation to the Circular Economy (CE), due to the direct relation with e-waste management. According to Fatimah, Govindan, Murniningsih, and Setiawan (2020), the use of resources from the 4th industrial revolution, such as IoT, information and communication technology (ICT), improves global waste management, creating value from resources, materials and energy to the circular economy of the city or country within it.

CE is a system of production and consumption. By that, it has the goal of efficiently maximizing the use of resources and waste (Chiappetta Jabbour et al., 2020), reducing the environmental impact, and increasing society's well-being.

As a global practice, urban mining offers the implementation of the CE with potential towards environmental savings (Arora et al., 2017), where e-waste is the object of work and the resource to be collected and treated to archive a greener and sustainable society, and therefore, a better economic status.

Urban mining saves energy and raw material, decreasing the CO2 emissions that create indicators to social and economic aspect for a greener city, by retrieving raw material from the e-waste, for example, by recycling 10kg of aluminum can save 90% of energy and prevents 20kg of CO2 emissions, when compared to the traditional mining (Kumar, Holuszko, & Espinosa, 2017).

Saving energy and reducing emissions and retrieving material such as gold, copper, aluminum, and steel also generates economic benefits for the region where a recycling unit is installed.

The adoption of innovative tools, such as an e-waste processing unit, can contribute to the CE and improve a region, especially when used at an optimal level. Adopting waste management practices has an impact on the social aspect of the city, generating jobs, improving the economy, and promoting new legislation that improves waste management, not considering that the social context of the waste is costly to policy goals and social effects (Gutberlet, 2018). It is known that in most countries, the selection, classification, and treatment of the waste is made by informal collectors, those units generate jobs through associations and cooperatives, transforming the waste management goals in benefit to the city economy, showing the importance of the presence of a recycling processing unit.

2.2 Sustainability indicators

The Smart City Index (SCI) generates a ranking of the most sustainable cities, considering three major areas, where the citizens' perception of a need to improve and current well-being is measured by the survey. They indicate Priority areas to be improved, by selecting the status of 15 indicators in a survey, that measures the five most important facts to be solved in their city.

The indicators measure the level and presence of affordable housing, fulfilling employment, unemployment, health services, basic amenities, school education, air pollution, road congestion, green space, public transport, recycling, security, city engagement, social mobility and corruption (Institute for Management

Development, 2020).

The SCI also measure the willingness of the citizens to participate in the growth of the city, as their attitude, where they allow the use of personal data to improve traffic congestion, increase the non-cash transaction on a daily basis, allowing the facial mapping to improve safety, lowering the crimes, and finally, it is measured the perception of trust on the government by the availability of information online.

The third area is based on perception and level of satisfaction with the structure and technologies available in the city regarding health & safety, mobility, and social activities, like green spaces and culture, opportunities, and governance.

The waste management is included on health & safety criteria by the item of “recycling services are satisfactory,” this shows the recycling in overall waste, not only for e-waste.

Other aspects should be included for the e-waste management, like mobility and opportunities, considering the traffic quality and the availability of service of collection and treatment of e-waste. The quality of the traffic shows the flow of waste to optimize the collection and reduction of CO_2 emission by the trucks that transport it.

2.3 Technological indicators

The Global E-waste Monitor 2020 is a report that discusses and presents the quantities flow and the circular economy potential for the field. It represents a study to aggregate data from the countries' e-waste generation, collection, and production of electronic equipment added to the country market (Forti et al., 2020).

The Global e-waste monitor database allows the understanding of the actual status of the WEEE, presenting information and technology that enable the analysis in many contexts.

The importance of adopting efficient e-waste management and the representativity of the technology needed to retrieve as much raw material as possible from the e-waste is part of the actions to improve the ecosystem for futures action, listed on The Global e-waste monitor 2020.

It covers the electrical and electronic equipment (EEE) in a group of organizations, temperature exchange equipment, screens and monitors, lamps, large equipment, small equipment and small it and telecommunication equipment, this last one representing the majority in the potential of retrieving raw material and rare metals.

The waste and recyclable materials often travel long distances, causing an issue due to energy use, increasing air pollution, traffic, and noise (Gutberlet, 2018). In fact, not only the transportation but the exportation of waste, especially e-waste via transcontinental shipment, to developing countries from the developed ones, turns the need for a processing unit installation in a city into an urgent matter of need; it is a technology to fight the harmful actions against the environment.

Usually, the exported e-waste sent to developing countries are poor in rare metals, those rare metals are extracted from efficient recycling e-waste units installed in developed countries, the exportation is a crime but "*has become an institutionalized practice among certain corporations that pollute, dump toxic waste and make environmental crime victims of various global minorities*" (Simon, 2016), to prevent this, it is important to understand the need of innovative technologies installation withing the recycling e-waste processing units as part of critical investments.

One example of innovation and technology applied to the e-waste management is the South Korean Eco-Assurance System that has the goal to minimize the load of the waste by implementing systematic management (Rhee, 2016), from a network system, the three responsible for the collection in Korea, the government, producers and recycling companies can exchange data about the collections made by them, determining the numbers of collectors nationally wide to implement the reverse route system.

It must be considered that technologies used in Smart Cities will become e-waste in the near future, and acting now can prevent the accumulation of increasing e-waste generation. The main part of technology and the related waste generation must be listed and analyzed to help decision-making.

2.4 Legal indicators

Legislation is one of the most important indicators to understand efficient e-waste management and its potentials. Nowadays many countries have specific e-waste policies to respect the Basel Convention and reach the desired Sustainable Development Goals.

Unfortunately, the installation of a high-tech e-waste processing unit in developing countries is very difficult. To properly manage e-waste (collecting, storage, recycling, disposing), many developing countries must develop low-cost recycling procedures and should implement laws and regulation, nowadays, the developing countries legislation are not very effective (Hossain, Al-Hamadani, & Rahman, 2015), but represents the intention to improve the overall system.

More effective legislation applied globally is usually related to the Extended Producer Responsibility (EPR) principles, showing efficiency to prevent pollution and waste generation, being implemented in most industrialized countries.

EPR is defined as "*is a policy approach under which producers are given a significant responsibility – financial and/or physical – for the treatment or disposal of post-consumer products*" (OECD). It requires the manufacturer to take responsibility for its production until the final disposal, especially disposal, dismantling, and reuse (Cao et al., 2016), being the base for many directives, national policies, and presidential decrees around the world.

The power of legislation is represented by the e-waste collection status of a country; for example, Japan created the Home Appliance Recycling Law in 2001 and promoted a new framework, creating a new paradigm in waste management, producing technological innovation (Ogushi & Kandlikar, 2007).

Ogushi and Kandlikar (2007) show that the recovery rate for e-waste, after regulation reached around 69.5%, putting Japan as one model to be followed in terms of e-waste management, by making explicit the responsibility for taking back and recovering for different stakeholders and lowing the cost of recycling by adding the final consumer the fee for the recycling.

Legislation has many other examples to enforce its importance. The Europe WEEE directive in 2002 improved the collection and efficiency of recycling; in 2016, the targeting rate of recycling changed and improved the expected level to 45% of all electronic products put on the market, increased to 65% (Kumar et al., 2017) of all electronic products put on the market, as expected from the literature.

In Australia, the legislation in 2009, National Waste Policy, and in 2011, National Television and Computer Scheme, focused on increasing the recycling rate using international best practices (Kumar et al., 2017).

Many other countries used the legislation to improve the recycling rate, using the EPR as a base and implementing international best practices and its variants. China, for example, implemented the “Old for New” policy to provide subsides for consumers, it improved to 84.5% of the recycling rate (Cao et al., 2016), it was possible because the takeback system would give a financial advantage to the final consumer that was willing to dispose of the electronic device in formal recycling units, "*the "Old for New" policy motivated the generation of a more standardized and large-scale e-waste recycling industry in China*" (Cao et al., 2016).

China implemented many other legislations over time, improving and keeping the standards to other countries, such as Japan and countries in European Union. The Reverse Supply Chain and the Reverse logistics are not legislation but the standardized process of recovery, reuse, and disposal of the e-waste, that is used as a best practice in the world, sometimes considered as wider than the Reverse Logistics, the Reverse Supply Chain needs more investment and has a higher impact on the economic benefits and strategies to the companies (Doan, Amer, Lee, Phuc, & Dat, 2019).

All legislation found in the literature indicates that the investment in innovations and infrastructure was needed, leading to the interpretation that it is important to have focused legislation to predict the installation of an e-waste recycling unit.

Chapter 3 Methodology

3.1 Definition

The research proposed a two-level analysis regarding smart cities ranking position and cities´ e-waste management data. These levels are connected using artificial intelligence (A.I.) techniques for unsupervised data to find similarities within the data of the cities and the prediction outcome of a potential installation of an e-waste recycling processing unit in a city based on the same model.

To study this potential, highly certified and well-documented technologies frameworks were selected. To define the variables to be collected, previous literature shows indicators to economic, social, technological, sustainable, and legislative aspects.

To define the procedure, it was considered the Artificial Intelligence capability to analyze data similarities and potential of predictability to a given factor. To that, the designed framework adopted for that approach was clustering and prediction by artificial intelligence script written in R language.

The K-means clustering is an unsupervised machine learning algorithm with simple application and popular use; it makes inferences from inputted datasets using the vectors, not considering the known or labeled outcomes (Garbade, 2018).

The K-means is a partitioning-based algorithm that uses Euclidean distance between the points to determine the k-centroid of the clusters, defining the area of similar points by the given number of clusters to be plotted because its low need of computer resources and simplicity it is used largely from the community (Mehta, Bawa, & Singh, 2020).

The A.I. prediction model selected was the Generalized Liner Model due to regression and classification capability and its adhesion to the defined data.

After the definition, the methodology proposed flow was designed, resulting in the image, as shown in Figure 2. In general, after the definition of the scope and the literature review, the variables were determined, the data was acquired and prepared to be inputted on the K-means and Generalized Linear Model (GLM) algorithms to be compared, making it possible to reach a conclusion according to the hypotheses created.

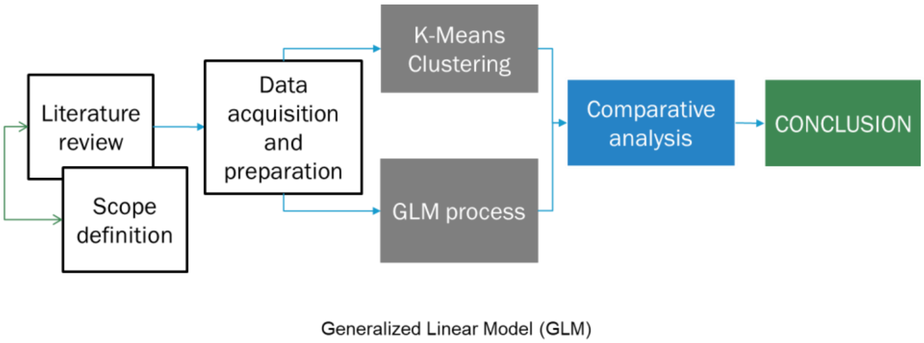

Figure 2. Methodology framework

The clustering process and the GLM process were defined to have a composition as a workflow to the framework, being adherent to the methodology, as shown in Figure 3 and Figure 4, respectively.

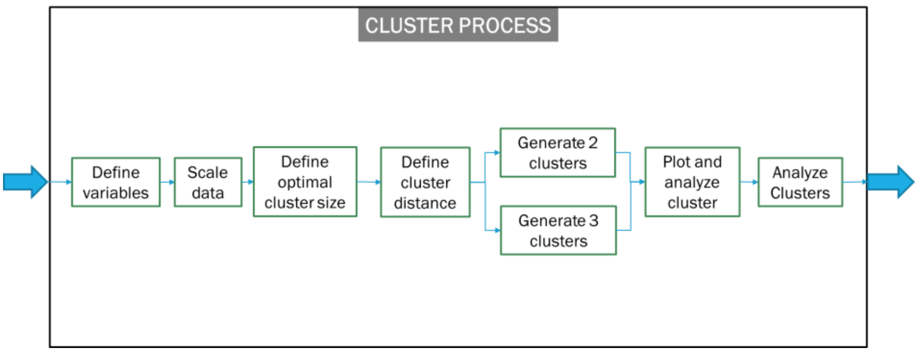

Figure 3. Cluster process framework

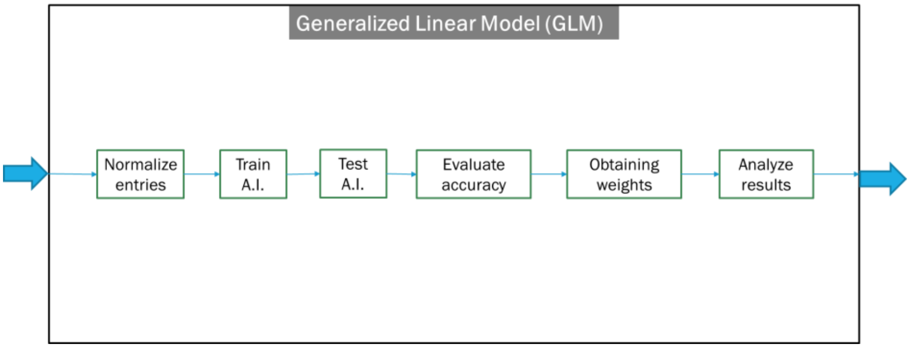

Figure 4. Generalized Linear Model process

3.2 Method analysis

To apply the method, it was needed to select the variables related to the literature review and defined scope; in that case, the variables chosen were related to the social and economic, sustainable, technological and legislation indicators from the databases available related to e-waste management and the smart cities.

The economic and social indicators selected were the Gross Domestic Product (GDP), population, the human development index, the size of the city, the quantity of (EEE) put on the city's market and the traffic quality.

For the sustainable indicators, it was selected the e-waste generated, the e-waste collected, and the smart city index of the city, to cover the relationship between the indicators and to allow the prediction.

The technical indicators led the research to select information related to the presence of seaports and the presence of e-waste recycling processing units, as described in the literature review its importance to the ICT level to an innovative status.

To the last indicators, the legislation available selected was the e-waste policy existence, an important item of analysis, also covered by the literature review.

The methodology variables availability was verified and adjusted, as shown in Tble 1, to the idea of Smart Cities' correlational numbers and standardized as quantitative values to get an easier analysis.

Table 1. Variables defined for the analysis

VARIABLES
Smart City Index
Population
Human Develop Index
Size in Km2
GDP in thousand USD
E-waste generated by kg
E-waster collected by kg
EEE put on the Market
Traffic quality

Existence of e-waste policy nationally
Quantity of seaports present on the city
Existence of e-waste recycling processing units

With the definition of the variables and aiming the answers to the research questions, a hypothesis was defined. It was based on the assumption that "*The e-waste processing unit brings innovation and circularity from the ICT to cities*."

Hypothesis: The city will be considered suitable for the installation of an e-waste recycling processing unit when:

- Smart City Index is high (1st, 2nd,…);
- GDP and HDI are high;
- E-waste generated and e-waste collected is high;
- E-waste policy is present.

3.3 Data gathering

The data gathering was possible after the variables definition, the Smart City Index and the Economic, from World Bank database, indicators were the first to be gathered, followed by the data of Global E-waste Monitor 2020, Our World in Data website, Global City Data, Google maps city area and the

SeaRates database for the Sea Port location.

Smart City Index 2020 Report has information from the smart cities, generated by applying a survey in 2019 for 109 cities defined as sustainable. The IMD World Competitiveness Center "assesses the perceptions of residents on issues related to structures and technology applications available to them in their city" (Center, 2020). The variables retrieved from the Smart City Index were the name of the cities, the city rank from 1st to 109th place, and the quality of the traffic, which is given from 0.0 to 1.0, depending on the citizens' evaluation of its perception, represented by a fraction, such as 0.921, for example. The traffic quality was selected by its high relevance demonstrated in the literature review. In the SCI Report, the cities are divided into groups by its Human Development Index (HDI), retrieved by country and associated with its cities from the United Nations Development Programme (UNDP). The HDI "was created to emphasize that people and their capabilities should be the ultimate criteria for assessing the development of a country, not economic growth alone" (Programme, 2020). Therefore, the economic and HDI growth will promote pollution up to the point that it will demand a higher sustainable city, decreasing the pollution, with an inverted U-shape form from the HDI and domestic waste treatment relationship. A higher level of sustainability will promote economic and HDI growth (Li & Xu, 2021).

The Global E-waste Monitor 2020 is a report collaborative from the Sustainable Cycles (SCYCLE) Programme, the United Nation University (UNU), the United Nations Institute for Training and Research (UNITAR), the International Telecommunication Union (ITU), and the International Solid Waste Association (ISWA) to monitor the quantities and flow of the e-waste globally, helping the decision-making towards a better society and to amplify the circular economy adoption.

Form The Global E-waste Monitor 2020 information was retrieved by country for the year of 2019, including the E-waste generated, the E-waste collected, the e-waste generated per capita, and the EEE put on the market.

The data from The Global E-waste Monitor 2020 is organized by country, the quantity of e-waste generated and EEE put on the market were gathered by kilogram (kg) per capita, then they were multiplied by the city's population to acquire the result in kg for each city.

The e-waste collected information is present as the rate of collection, based on the generation, also by country, to standardize to kg of e-waste, the rate of collection was multiplied by the population of the city, resulting in the final data of the quantity of kg of e-waste collected for each city.

The World Bank database is an open data platform that contains many aspects of countries in the world, especially about economic variables, the 2019 GDP and GDP per capita for each country was retrieved by accessing it.

To the defined variable, the GDP per capita of each country was multiplied by its cities' population; the result was the variable of GDP in thousands of dollars.

In Our World in Data, it was retrieved each city's population. The city's size was retrieved from other similar databases like Global City Data, published by the Greater London Authority (GLA), and many others like Google Earth, confirming the size in square kilometer.

From SeaRates, by DP World, it was possible to retrieve the presence of ports by searching each one of the 109 cities listed on the SCI. It represents the capability and the opportunity to receive the e-waste via sea, measuring its importance while having an e-waste recycling processing unit installed.

The data gathering procedure is summarized in Table 2, as follows.

Table 2. Data gathering

Data type	Source and weblink
Smart Cities indicators	Smart City Index https://www.imd.org/smart-city-observatory/smart-city-index/
Population	Our World in Data https://ourworldindata.org/
Human Develop Index	Human Development Report http://hdr.undp.org/en/data
Gross Domestic Product (GDP)	World Bank https://data.worldbank.org/indicator
Port location	Sea Rates https://www.searates.com/pt/
E-waste generation	Global E-waste Monitor http://ewastemonitor.info/
E-waster collected by kg	Global E-waste Monitor http://ewastemonitor.info/
Traffic quality	Smart City Index https://www.imd.org/smart-city-observatory/smart-city-index/
E-waste policy nationally	Global E-waste Monitor http://ewastemonitor.info/
Existence of e-waste	Science Direct search

recycling processing units	https://www.sciencedirect.com/ Open-source web references
Equipment put on the market	Global E-waste Monitor http://ewastemonitor.info/
City size	Global City Data & Google Earth https://data.london.gov.uk/dataset/global-city-data https://earth.google.com/web/

3.4 Data arrangement

All the data gathered were related to 2019, inserted on a CVS, verified and double-checked by specialists, adjusted to be understood by the RStudio and saved with the standard format for numbers, with a point as decimal separator and a comma as a thousand separator.

For the prediction on the artificial intelligence, it was configurated to receive the data as quantitative variables, with independent values, but with the prediction factor as the presence of the e-waste recycling processing unit. It became the dependable variable with a qualitative aspect.

Chapter 4 Analysis and Results

4.1 Initial procedures and environment

The dataset arranged was organized in the order of city name, SCI, city population, HDI, size in Km2, GDP of the city in thousand USD, E-waste generated on the city in Kg, e-waste collected on the city in Kg, EEE put on the market by kg, Level of traffic quality at the scale of 0.000 to 1.000, the existence of e-waste policy as 0 or 1, presence of seaport on the city, with the number of seaports in the city, 0, 1 or 2, and presence of e-waste recycling processing unit, identified and determined as 0 or 1, where 0 is not present, and 1 is considered present.

The file was saved in CSV, encoding UTF8, separated by semicolon and comma as decimal separator. The hardware used for the analysis was an HP Pavilion Laptop 15-cc5xx series, composed of processor i5 7200U 2 physical and 4 virtual cores, 16Gb RAM DDR4, SSD M.2 480Gb, Intel HD Graphic 620 integrated with 1Gb being used.

The software included Windows 10 pro, building 19042.985, RStudio 1.4.1106, using the R libraries dplyr, readr, e1071, caret, ggplot2 FactoMineR, factoextra, tidyverse and cluster, with the latest version available for the current RStudio.

4.2 K-means clustering analysis and results

According to Nazeer and Sebastian (2009), "*Clustering is the process of partitioning a given set of objects into disjoint clusters. This is done in such a way that objects in the same cluster are similar while objects belonging to different clusters differ considerably, with respect to their attributes.*"

The cluster analysis is a technique used to identify similarities and generate observations over the aggregated points providing insightful observations and considerations (Govender & Sivakumar, 2020).

The K-means analysis needs scaled data from numeric values; in that case, the first variable, the name of the city, was set as the rows' names, labeling the data in each row of the other variables.

The K-means clustering approach tried to understand the logic applied to the algorithm since it would use many variables and not only two to determine the points in a 2D graphic to be clustered.

To reduce the variables to two dimensions, the K-means uses the Principal Component Analysis (PCA), which is a statistical technique for unsupervised dimension reduction, closely related to the unsupervised learning, creating the best low-dimensional linear approximation of the data (Ding & He, 2004).

According to Lopes (2017), the PCA algorithm "use the concepts of variance matrix, covariance matrix, eigenvector and eigenvalues pairs to perform PCA, providing a set of eigenvectors and its respectively eigenvalues as a result."

To better analyze de data acquired, it was made two approaches, first one was the removal of the cities without all information, from that procedure, 87 cities were remaining, the main reason was that the variables e-waste collected and the EEE put on the market were missing on the 22 removed cities.

The second approach considered the lack of information for the e-waste collected and EEE put on the market, once it is needed to test and check the centroids[2], to improve the cluster accuracy by giving it more data, generating more data points (Nazeer & Sebastian, 2009).

Only one city was removed from the study for the lack of e-waste management information. Taipei City does not have any data related to The Global E-waste Monitor 2020, and it was removed from the analysis.

[2]A centroid is a data point (imaginary or real) at the center of a cluster (Söder, 2008).

After scaling the data, the PCA was generated to analyze the eigenvalue[3] and the variance of each variable, the difference of the variance and the eigenvalue the lower bounds for K-means (Ding & He, 2004).

4.2.1 Full set clustering analysis and results

To analyze the data in its full variables selected, the 87 cities were applied to the nominated "Full set," it is the first approach, where all cities have data in all variables.

To generate the dimensions, the K-means clusters algorithm used the PCA technique. It is one of the most used dimension reduction techniques (Fodor, 2002), and it generates the representations of the variance within the variables; the most representative distinguish features are more considered than the similarities, generating the data representation (Lopes, 2017).

The higher variance creates the data representation, and for the Full set, the percentage of this variance created, as shown in Figure 5, two dimensions with 36.9% and 20.5% of all data representation, the more the variable is distinct in

[3]Eigenvalues are a special set of scalars associated with a linear system of equations (i.e., a matrix equation) that are sometimes also known as characteristic roots, characteristic values, proper values, or latent roots. (Weisstein, 2021)

data from other (Marsboom, Vrebos, Staes, & Meire, 2018), more it will represent the dimension on PCA. The variance on the variables creates dimensions 1 and 2 to be used on the graph and the cluster.

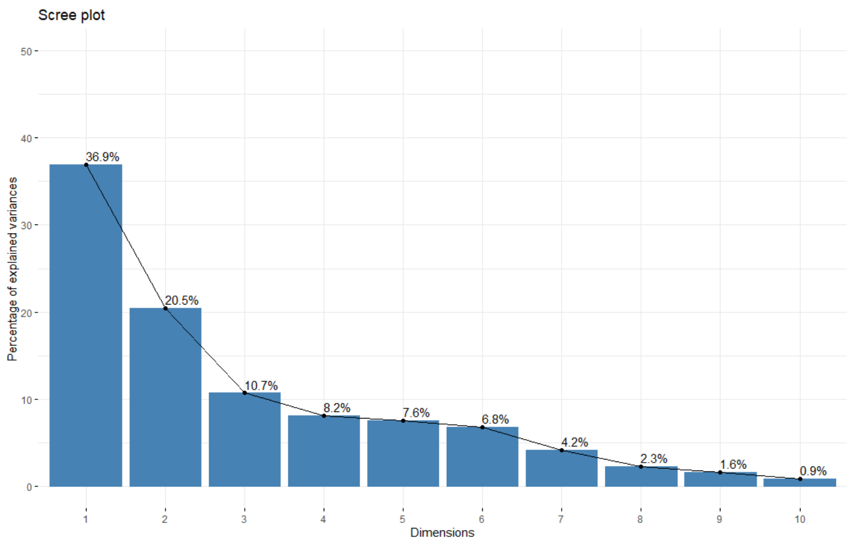

Figure 5. Percentage of variances – 87 cities

The quality of that variance is expressed on the representation graph, which demonstrates the influence of each variable to create each dimension, the bigger the blue circle, the bigger the representation, as can be seen in Figure 6.

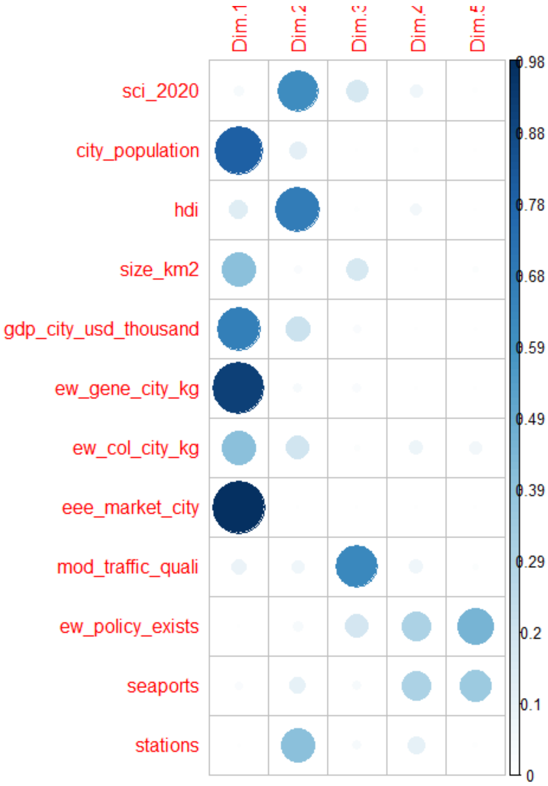

Figure 6. Quality of representation – 87 cities

From Figure 6, it is possible to understand that dimension 1 has a higher representation from the EEE put on the market (*eee_market_city*), followed by the e-waste generated (*ew_gene_city_kg*), the city population (*city_population*) and the GDP of the city (*gdp_city_usd_thousand*).

Dimension 2 is mainly represented by the variance of the variables HDI, SCI (*sci_2020*), and the presence of e-waste recycling processing units (*stations*), in that order.

To understand the position of each point represented on the cluster by the two newly created dimensions, it is important to observe the directional representation of each variable, where dimension 1 is plotted on the X-axis and dimension 2 on the Y-axis, as shown in Figure 7.

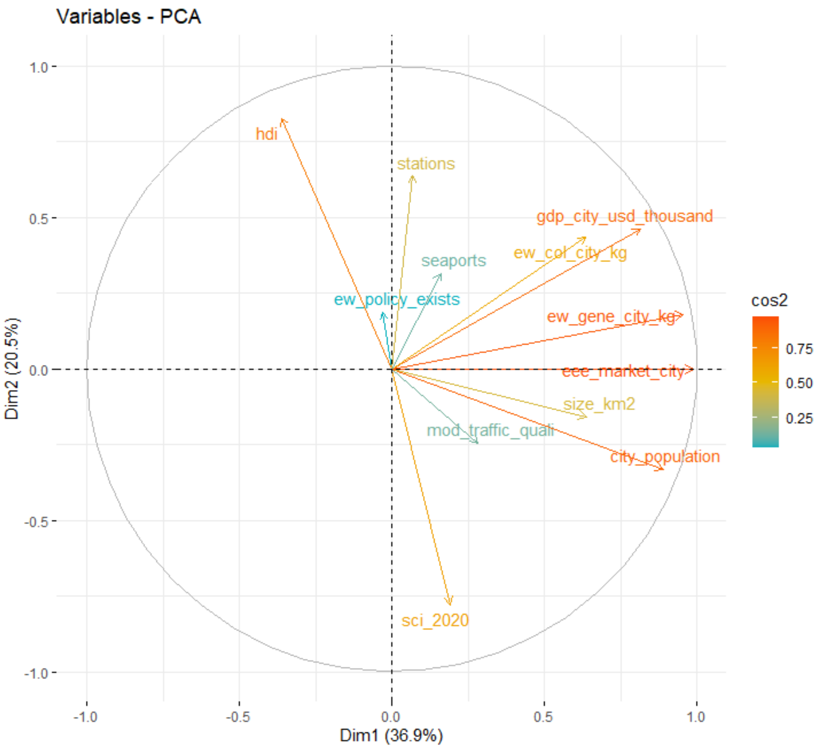

Figure 7. Variable in PCA representation – 87 cities

Each value of the variable "pulls" the point in its direction, creating the approximation by similarity desired on the study, and therefore generating the cluster by the proximity of each point to the centroid defined by the selected number of clusters.

Confirming the quality of the representation, the variable in PCA representation will force the coordinates X and Y for each city in the cluster to be plotted. After the PCA analysis, it is needed to generate the cluster from the data with the same aspects.

For the Full set, the optimal number of clusters was calculated to generate the graphical observation. In Figure 8, it is possible to see the result of the function that calculates the optimal number of the cluster as 2, given the dataset and the method of the silhouette of it.

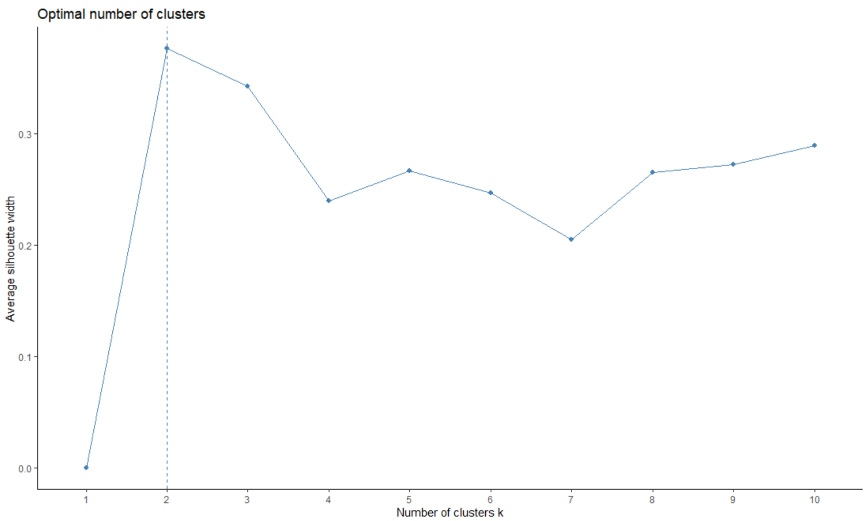

Figure 8. Optimal number of clusters – 87 cities

To compare the optimal number of clusters of 2, it was also generated a graph with 3 clusters; this can help to understand the difference from both images and increases the insight on the analysis, the technique of observing different cluster numbers from the same data is vastly used as can be observed by the study of Ünlü and Xanthopoulos (2019).

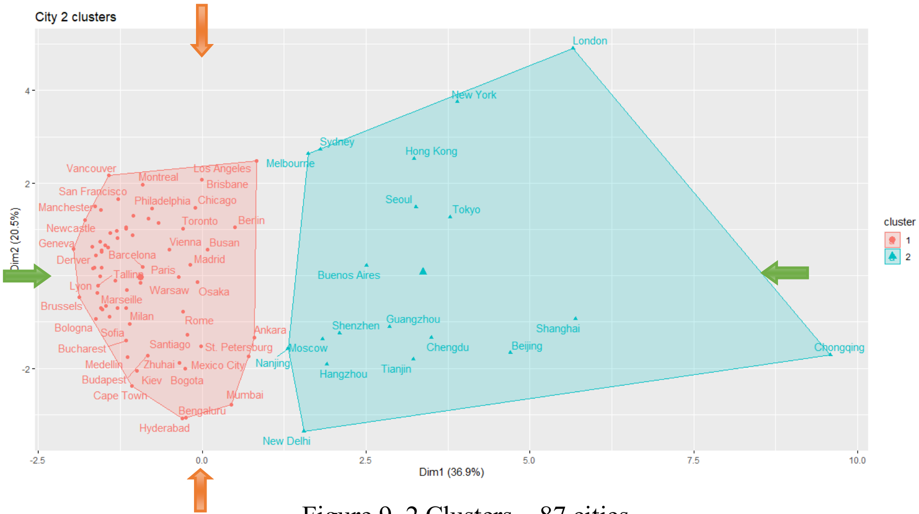

Figure 9. 2 Clusters – 87 cities

The Figure 9 graph shows the representation of 2 clusters of similar cities, considering the variables reduced to two dimensions. On cluster number 2, are cities mostly from Asia countries like Chongqing, Shanghai, Hong Kong, Seoul, Tokyo, Shenzhen, Guangzhou, Chengdu, Beijing, Tianjin, and Nanjing with similarities with other world cities, such as New Delhi, Moscow, London, New

York, Sydney, Melbourne and Buenos Aires, considering the similarities on values of the GDP, E-waste generation, EEE put on the market and city population.

Cluster number 1 has the other cities with similar variables, showing more cities with the same characteristics but on different continents.

The green and orange arrows marking the point 0 of scaled variables, on the upper side of the green arrows are the cities with higher HDI, lower SCI (Higher ranking 1º, 2º, etc.), and presence of e-waste recycling processing units, and on the right side of the orange arrows are the cities with a higher quantity of EEE put on the market, e-waste generated, city population and GDP.

The representation of the clusters indicates that for cities with higher economic power, e-waste generated and collected, as much as population and spatial size would be more likely to be an optimal place of an e-waste recycling unit.

The clusters are divided mainly by Dimension 1, putting cities with higher EEE put on the market, the e-waste generation, the city population, and the GDP into cluster 2 and with lower levels of those into cluster 1.

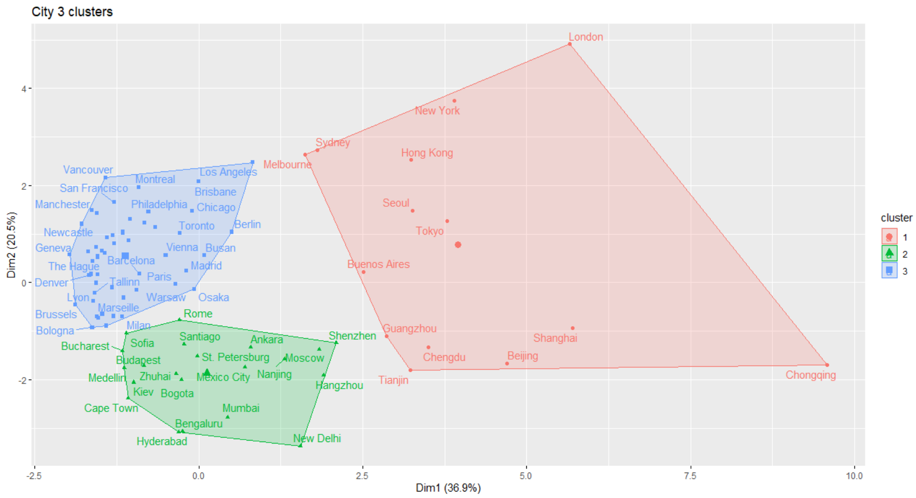

Figure 10. 3 Clusters – 87 cities

Using the same data but generating a graph with 3 clusters defined, it is showed a division on the previous cluster number 1 of Figure 9. The new cluster number 2, represented in Figure 10, but the similarity on many cities with higher rank of SCI and higher HDI, but still on the group of lower economic power, e-waste generation and collection, and city population and size.

To consider that cluster 1, in Figure 10, is now represented by fewer cities than before, we can relate this result to the same higher quantity of EEE put on the market, e-waste generated, city population and GDP, but with a higher level of difficulty to analyzing, ensuring that the K-means with 2 clusters is really the optimal value, as it is more classificatory.

4.2.2 Full cities dataset clustering analysis and results

One main concern of the previous analysis was that the 109 smart cities were reduced to 87, representing ~79.82% of the data on the dataset, caused by the lack of data of some cities on the variables of e-waste collection and quantity of EEE put on the market, as already explained previously on this study.

It is clear on the PCA analysis that both variables are highly representatives, especially the EEE put on market. Still, to compare the results, those two variables were removed from the dataset, keeping the e-waste generation and e-waste policy as representations of the e-waste management. It was again analyzed, considering the possibility of getting more points to increase the accuracy.

The first result is the percentage of the variance of variables reduction; the two dimensions changed to 29.1% and 24% of the overall variables, decreasing the data representativeness from the previous model, as shown in Figure 11.

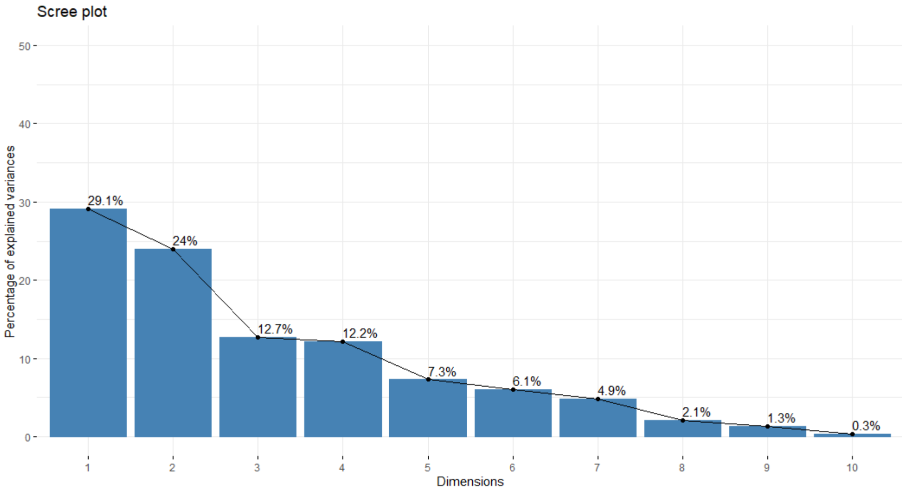

Figure 11. Percentage of variance - 108 cities

The removal of the variables and the increase of the numbers of the city decreased the variance to dimension 1. They increased it to dimension 2, denoting that the sum of all data variance representation is now 53.1%. In comparison, in the first variance, it was 57.4%; in other words, the two dimensions representation for 10 variables is 53.1% of all data (Nistrup, 2019).

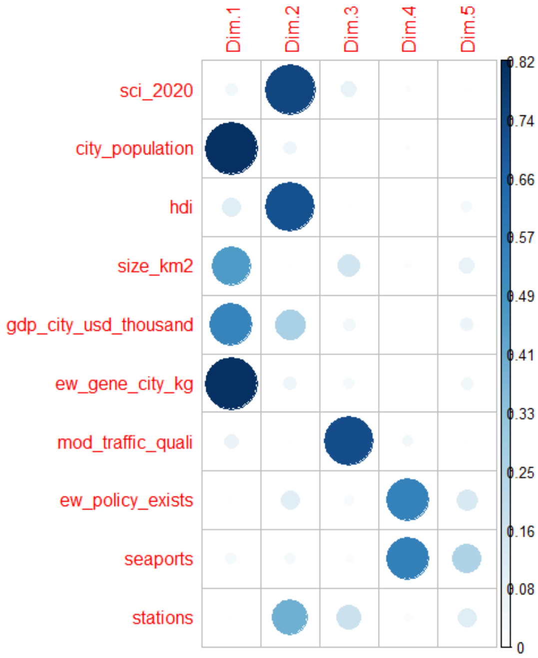

Figure 12. Quality of representation - 108 cities

The quality of representation with 108 cities, presented in figure 12, shows that now, the city population, the e-waste generated, the GDP and the size of the city are the determining dimension 1, without the EEE put on the market, the population is the most variant variable, the e-waste generation keeps as second, as the GDP the third, but now the size of the city is shown.

The SCI, the HDI, and the e-waste recycling processing unit (stations) continue defining dimension 2, but with the switch of importance between the SCI and the HDI.

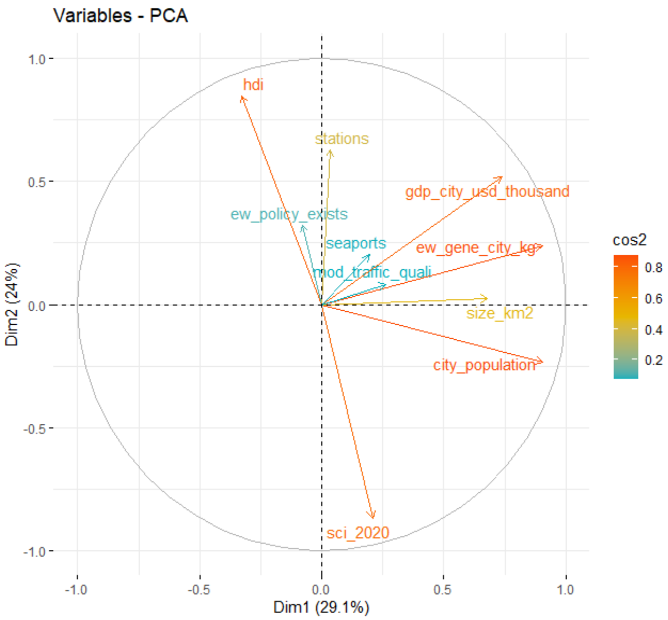

Figure 13. Variable in PCA representation - 108 cities

The variables in PCA representation, presented in figure 13, represented some loss of the force for seaports and traffic quality, shifting the second one to force the points over the two dimensions upwards instead of downwards.

Another observation is that for the presence of an e-waste recycling unit (stations), its representation increased, creating the perspective that this would be a better model for prediction than the first one.

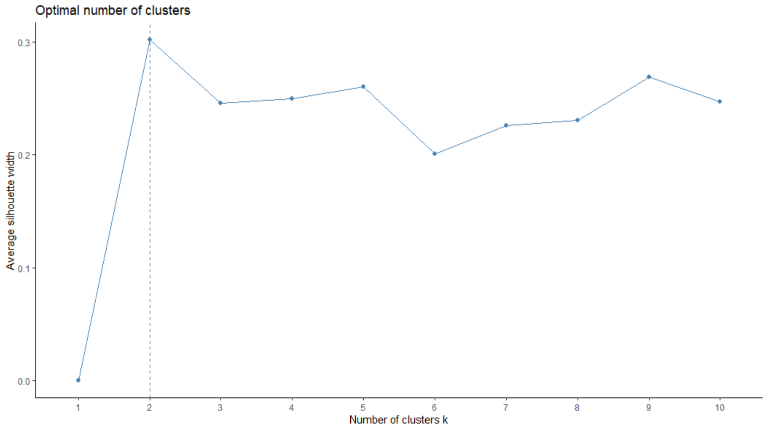

Figure 14. Optimal number of clusters - 108 cities

The optimal number of clusters continues 2, Figure 14, representing the same similarity as before.

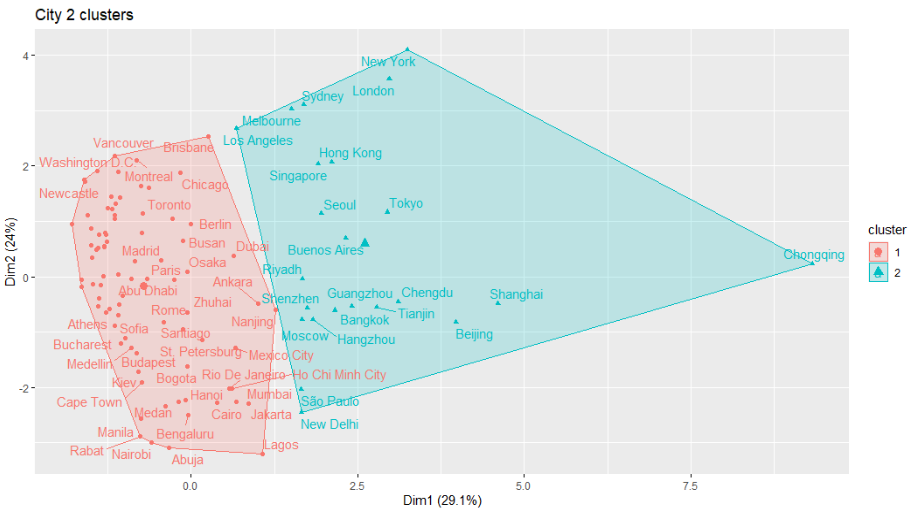

Figure 15. 2 Clusters - 108 cities

In Figure 15, the shapes of the clusters are similar. Still, the proximity of them on the second model is higher; in fact, the group of cities did not vary; the observation for that second approach is the inclusion of São Paulo, Los Angeles and Singapore at the cluster number 2, at the first model, those cities were excluded.

From the first and second approach, they are valid and similar, which allows the use of both for the prediction model, but the second approach can be considered less precise since the data variance representation is lower (Lopes, 2017).

4.3 Generalized Linear Model analysis and results

The Generalized Linear Model (GLM) is a statistical analysis with broad use, it generalizes the multiple correlation coefficient, is numerically simpler to interpret (Zheng & Agresti, 2000), doesn't need to scale the data, and can be used in unnormalized data, which represents the use of all possible data acquired for the Smart Cities, and the dataset available, arranged to the study.

The procedure for prediction was developed in RStudio, all data models of the K-means clustering analysis were used to keep the base of comparison, but for the prediction, it is mandatory to select an outcome, the presence of an e-waste recycling processing unit (stations) was selected as the factor to predict the installation or not of it in a city.

Additionally to the clustering procedure, the acquired data was divided into two sets, the training, and the test, using the createDataPartition function that "*can be used to create stratified random splits of a data set*" (Kuhn, 2008), with a division of 75% and 25% of the data, respectively. This function determines the training, and the test sets, creating a partition from the acquired data; it permits to create sets with data similarity, ensuring the reproducibility of the prediction. The use of GLM allows the analysis "*to tackle a wider range of data with different types of response variables*" (Faraway, 2016), which contributes to the dataset arranged and improves the possibility of success.

4.3.1 Full set prediction

The first approach, with 87 smart cities, was also used to the prediction algorithm. It generated 66 samples of linearity and 11 predictors for analyzing the outcome of the classes 0 and 1, where 0 is the absence of stations and 1 is the presence.

The first approach, with 12 variables, returned an accuracy of 90.48% for the prediction of presence or not of the stations, as shown in figure 15.

```
Confusion Matrix and Statistics

test_pred  0  1
        0 17  2
        1  0  2

          Accuracy : 0.9048
```

Figure 16. Prediction result - 87 cities

The prediction also shows the linearity of the data distribution for the cities in some graphics, as shown in Figure 17, they provide more validity to the approach by the observation of the prediction algorithm (red line) and the actual data (black dots) of the smart cities.

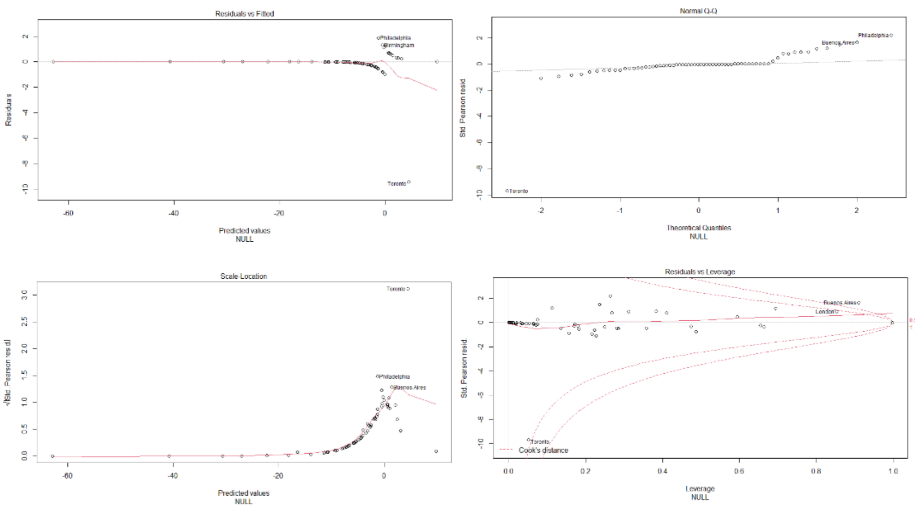

Figure 17. Prediction graphs – 87 cities

The model also outputted the importance of each variable for the prediction, the most important variables to predict were the GPD, E-waste collected, city population, EEE put on the market and the SCI, all of them weighing above 1.0, as shown in Figure 18.

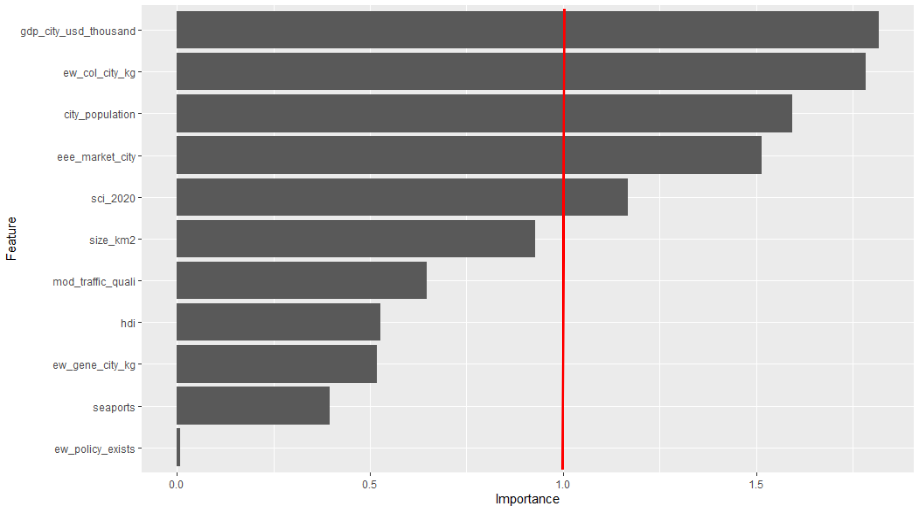

Figure 18. Variable importance to the prediction – 87 cities

According to Figure 18, the city size, traffic quality, HDI, e-waste generation, seaports, and e-waste policy are also presented and under-qualified indicators (< 1.0).

4.3.2 Full cities prediction

For the second approach, the same procedure was applied. The data was divided, trained, and tested. It generated 81 samples with 9 predictors to the same classes 0 or 1, into the presence of an e-waste recycling processing unit (stations) or not.

The prediction of the approach, without the e-waste collected and the EEE put on the market, generated an accuracy of 92.31% when determining the e-waste recycling processing unit (stations) presence, as shown in Figure 19.

```
Confusion Matrix and Statistics

test_pred  0  1
        0 21  1
        1  1  3

               Accuracy : 0.9231
```

Figure 19. Prediction result - 108 cities

That accuracy can denote a better approach for predicting the installation of the considering more cities and removing the collection and EEE production.

The linear graphics of the prediction, shown in Figure 20, shows lines as the previous one, but with some level of degradation of the linearity, with more outlines, but with the validity of the approach by the prediction (red line) and the actual data (black dots) of the smart cities, in the same way of the previous dataset.

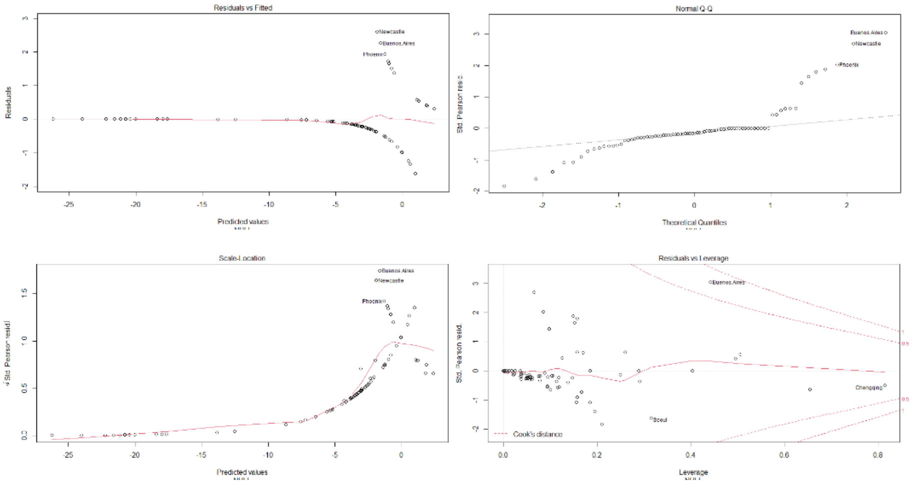

Figure 20. Prediction graphs - 108 cities

The variable importance of the prediction in the second approach, with 108 cities, showed that only the traffic quality and the SCI were higher than 1.0 for the weight shown in Figure 21. It shows the tendency to consider smart cities as more suitable for an e-waste recycling processing unit installation when the traffic quality is higher.

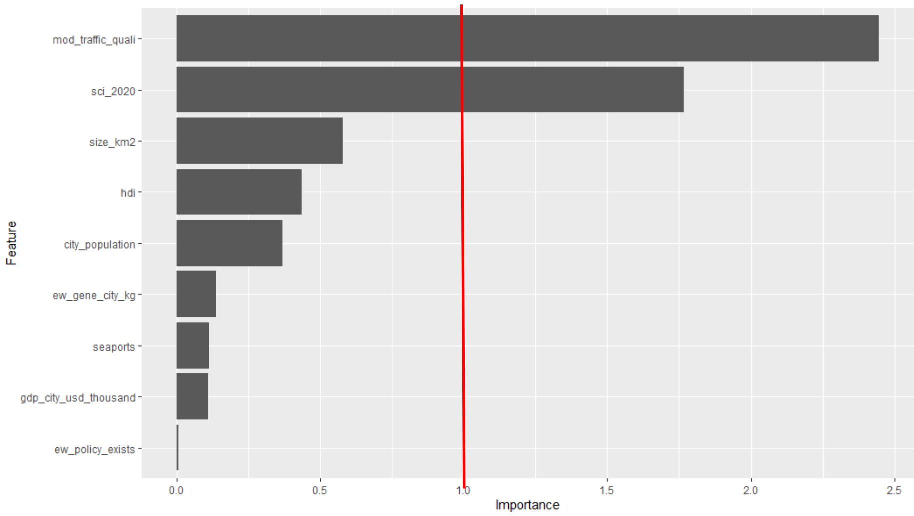

Figure 21. Variable importance to the prediction - 108 cities

4.3.3 Comparison from models results

One of the most important results for the GLM used is the observation of the response of the prediction instead of the possible applied rank (Zheng & Agresti, 2000). It is driven to the accuracy of the prediction, on Table 3, and it is possible to see the relationship between approach 1, with all data, and approach 2, where the e-waste collection and EEE put on the market data are not considered.

Table 3. Confusion Matrix and Statistics comparison

Confusion Matrix and Statistics							
Prediction test		**0**	**1**			**0**	**1**
	0	17	2		**0**	21	1
	1	0	2		**1**	1	3
Prediction factors	**All Data**				**No EW Collected No EEE no Market**		
Accuracy	90.48%				**92.31%**		
95% CI	(0.6962, 0.9883)				(0.7487, 0.9905)		
No Information Rate	**80.95%**				84.62%		
P-Value [ACC > NIR]	20.77%				21.40%		
Kappa	61.82%				70.45%		
Mcnemar's Test P-Value	47.95%				100.00%		
Sensitivity	**100.00%**				95.45%		
Specificity	50.00%				**75.00%**		
Pos Pred Value	89.47%				**95.45%**		
Neg Pred Value	**100.00%**				75.00%		
Prevalence	80.95%				84.62%		
Detection Rate	**80.95%**				80.77%		
Detection Prevalence	**90.48%**				84.62%		
Balanced Accuracy	75.00%				**85.23%**		
Positive' Class	**0**				**0**		

Table 3 shows some relevant factors to be explained, not only the accuracy but the No Information Rate, Sensitivity, Specificity, the Positive prediction value (Pos Pred Value), the Negative prediction Value (Neg Pred Value), Detection Rate, Detection Prevalence and the Balanced Accuracy has some insightful meaning.

First, the No Information Rate shows the rate of values “No” represents on the prediction, showing a lower “No,” where there is NO e-waste recycling unit installed, intends to be denoting a distribution with more perspective of E-waste recycling unit installation, and therefore, a better approach to the study (HRanalytics, 2017).

Considering that the prediction algorithm recognized the 0 as positive cases, the data from approach 1 showed a higher negative prediction value, which is more relevant to the installation of the E-waste recycling processing unit, and a higher Specificity that denotes the also the higher prediction of negative values, class 1, to the installation of the stations (HRanalytics, 2017).

The Detection Rate shows a higher value for approach 1; it represents the rate of detection at all data, while the Detection Prevalence is all actual positives rate in the whole data, also higher for approach 1.

One more fact is that the Balanced Accuracy is higher for approach 2, without the E-waste Collection, the quantity of EEE on the market, and with 108 Smart Cities data (HRanalytics, 2017).

The other variables have some representation, such as the McNemar's Test P-Value, that "*In statistics, McNemar's test is a non-parametric method used on nominal data to determine whether the row and column marginal frequencies are equal*" (Sun & Yang, 2008) shows the marginal frequency of the data, representing the variables independency.

Chapter 5 Conclusion

E-waste generation is a global problem that has increased exponentially from the last years due to the 4th industrial revolution and all technology's development, innovation, and consumption. Adding more ICT into the e-waste management has the power of implementing the capacity of treating it, giving a proper disposal, and allowing the implementation of Circle Economy with higher effectiveness.

In that direction, one possible action is the installation of E-waste Recycling Processing Unit in a City; this understanding led the research to propose the modeling of data from Smart City Index, E-waste management features, economic, social, legislation, and technological aspects to analyze that possibility and create a viable model to predict the optimal city for installation. By using two approaches as methodology, the classification by clustering using K-means algorithm and prediction by using the Caret algorithm and the Generalized Linear Model, the research acquired some interesting result that gives some highlights and insightful perception of the data used.

By questioning "What are the key indicators of smart cities that makes them satisfactory location of e-waste management facilities?" and if "Is there any similarity between the classification and prediction models that helps to identify the correlations between the e-waste management and the Smart City Index?", the research was modeled based on the Smart City Index, the GDP of the city, the size, the generation and collection of e-waste on the city, the population, the HDI, the traffic quality, and the existence of policies related specifically to e-waste management, the presence of seaports of the city and the presence of e-waste recycling facilities.

From that data, it was possible to generate the K-means cluster analysis and to create a prediction model using the Generalized Linear Model as a method; the first approach showed that to use all variables, the numbers of 109 cities would drop to 87, mostly because of the lack of information for E-waste Collection and EEE put on the market, what led the research to analyze the data with two possibilities, the 87 cities with all variables and 108 cities without considering the E-waste Collection and EEE put on the market.

The first analysis results showed that 87 cities, divided into two clusters, as indicated on optimal cluster number, are separated by Dimension 1, with 36.9% representation of the data, and with the quantity of EEE put on the market, e-waste generated, city population and GDP as determinant variables to the cluster.

The variance on that analysis showed a higher representativity of what is more distinguished inside the data, increasing the similarity by that data variance (Lopes, 2017).

Comparing to the second analysis, the 108 cities on two clusters showed that the distribution was similar, but dimension 1 was created with a 29.1% representation of the data, with the city population, the e-waste generated, the GDP and the size of the city as the variables to difference them. It is important to consider that Dimension 2 on both datasets was mostly represented by the SCI, HDI and traffic quality.

In fact, looking at those aspects represented by the variables, the set of cities did not have much difference, but the representation of the first analysis shows two variables related to e-waste management, while the second one only one, and not listed as the most representative characteristics, it reached the second most representative.

That difference led us to conclude that the first cluster would represent much more the groups of smart cities related to the e-waste management and capability than the second one, denoting those cities with higher GDP, e-waste generation and collection, population, and size would be more likely to be prepared to evolute on the next step of the e-waste management, increasing, or installing a new one, e-waste recycling unit.

The research's question can be solved with the indication of the variables in order of importance: Quantity of EEE put on the market; E-waste generated; City Population; and GDP, putting them as the key indicators to reach that of what needs to improve, to increase the e-waste recycling rate, by the adoption and implementation of the e-waste recycling unit.

The quantity of EEE put on the market and e-waste generated are related to the size of population and its GDP, characterizing one cluster as cities richer, more populated with the capacity to produce EEE and generate e-waste to be considered the optimal place to install one E-waste recycling processing unit, by this method.

The second method was the prediction, using the same two variances of the data, with 87 cities and all data on the variables, and 108 cities cutting out the collection and the quantity of EEE put on the city's market.

Despite the higher accuracy of 92.31%, the approach with 108 cities, without the collection and EEE put on to the market, showed that it considers the traffic quality and the SCI more important than all other variables; the analysis highlighted that more information could be considered better for the prediction. In comparison to the dataset with 87 cities and all the variables, the characteristic represents the higher impact from GPD, E-waste collected, city population, EEE put on the market and the SCI, which have more relations with the representation of e-waste management than the first one.

The implications of adopting a prediction model for a city, such as the one presented here, considering the SCI as a standard, can drive the creation of a framework to improve the city aspects to reach the goal of installing an e-waste recycling unit.

The decision-making on investment for the city e-waste management, observing the results discussed, should consider the key indicators as steps of evolution towards a smarter and cleaner city.

The EEE market, the control and collection of the e-waste production emphasize the need for investment in the local industry of EEE and the need for policies to increase the GDP by implementing and supporting the Circular Economy.

One conclusion to the prediction method, looking at the explained Detection Rate, the Detection Prevalence, and ultimately by the No Information Rate, is that more variables, more data, with more relation to the e-waste management would be more precise to determine the city characteristics of the optimal place of installing one E-waste recycling processing unit, with possibility prediction in many other aspects, like policy, prediction of collection, and et.

Both methods, the cluster and the prediction, converge in the same direction; more variables related to the e-waste is more precise and more representative, showing that: Quantity of EEE put on the market, Population, GDP, E-waste collection, and SCI are the most important index to reach a more appropriated e-waste management, while determining the E-waste recycling processing unit installation.

The research hypothesis of considering one city more suitable for the installation of an e-waste recycling processing unit can be confirmed to SCI, GDP, HDI, E-waste generated and collected, but need more discussion when approaching the presence of E-waste related policies. The variables confirmed on the hypothesis need to be put on a rank, as shown by the results.

The limitations found in the research were the quantity of data for the e-waste management. More data would be needed to compare the variables for all Smart City, and more Smart Cities Index would also contribute to a better result.

The contribution of the research is the discussion of the model with the SCI and its characteristics to be valid as a standard goal to the e-waste management-related subjects.

Smart Cities can also contribute more to the E-waste management and collaborate to solve the E-waste generation problem faster by its features of innovation and efficiency, represented by the data correlation.

The following research could approach other aspects of the prediction to the same amount of data, also, it is needed more research and discussion about the quantity of Smart Cities on the Smart City Index, as well as the data from the E-waste, to understand better the World Demand to be a cleaner place.

Finally, research on the direction of the impact of the policies and its features, contextualized as numerical factors, would help the next steps of the e-waste management, not only discussing the observation of regional impact but also creating a possible standard globally that can be applied to future Smart Cities development.

Bibliography

Abbondanza, M. N. M., & Souza, R. G. (2019). Estimating the generation of household e-waste in municipalities using primary data from surveys: A case study of Sao Jose dos Campos, Brazil. *Waste Manag, 85*, 374-384. doi:10.1016/j.wasman.2018.12.040

Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., & Airaksinen, M. (2017). What are the differences between sustainable and smart cities? *Cities, 60*, 234-245. doi:https://doi.org/10.1016/j.cities.2016.09.009

Arora, R., Paterok, K., Banerjee, A., & Saluja, M. S. (2017). Potential and relevance of urban mining in the context of sustainable cities. *IIMB Management Review, 29*(3), 210-224. doi:10.1016/j.iimb.2017.06.001

Benites, A. J., & Simões, A. F. (2021). Assessing the urban sustainable development strategy: An application of a smart city services sustainability taxonomy. *Ecological Indicators, 127*. doi:10.1016/j.ecolind.2021.107734

DECRETO Nº 10.240, DE 12 DE FEVEREIRO DE 2020, 10.240 C.F.R. § 1 (2020).

Cao, J., Lu, B., Chen, Y., Zhang, X., Zhai, G., Zhou, G., . . . Schnoor, J. L. (2016). Extended producer responsibility system in China improves e-waste recycling: Government policies, enterprise, and public awareness. *Renewable and Sustainable Energy Reviews, 62*, 882-894. doi:10.1016/j.rser.2016.04.078

Center, I. W. C. (2020). *IMD-SUTD Smart City Index Report*. Retrieved from https://www.imd.org/smart-city-observatory/smart-city-index/

Cheela, V. R. S., Ranjan, V. P., Goel, S., John, M., & Dubey, B. (2021). Pathways to sustainable waste management in Indian Smart Cities. *Journal of Urban Management*. doi:10.1016/j.jum.2021.05.002

Chiappetta Jabbour, C. J., Seuring, S., Lopes de Sousa Jabbour, A. B., Jugend, D., De Camargo Fiorini, P., Latan, H., & Izeppi, W. C. (2020). Stakeholders, innovative business models for the circular economy and sustainable performance of firms in an emerging economy facing institutional voids. *J Environ Manage, 264*, 110416. doi:10.1016/j.jenvman.2020.110416

Choksi, S. (2001). The Basel Convention on the control of transboundary movements of hazardous wastes and their disposal: 1999 Protocol on Liability and Compensation. *Ecology LQ, 28*, 509.

de Oliveira, C. R., Bernardes, A. M., & Gerbase, A. E. (2012). Collection and recycling of electronic scrap: a worldwide overview and comparison with the Brazilian situation. *Waste Manag, 32*(8), 1592-1610. doi:10.1016/j.wasman.2012.04.003

Ding, C., & He, X. (2004). *K-means clustering via principal component analysis.* Paper presented at the Proceedings of the twenty-first international conference on Machine learning.

Doan, L. T. T., Amer, Y., Lee, S.-H., & Phuc, P. N. K. (2019). Strategies for E-Waste Management: A Literature Review. *International Journal of Energy and Environmental Engineering, 13*(3), 157-162.

Doan, L. T. T., Amer, Y., Lee, S.-H., Phuc, P. N. K., & Dat, L. Q. (2019). E-Waste Reverse Supply Chain: A Review and Future Perspectives. *Applied Sciences, 9*(23). doi:10.3390/app9235195

Dutta, D., & Goel, S. (2021). Understanding the gap between formal and informal e-waste recycling facilities in India. *Waste Manag, 125*, 163-171. doi:10.1016/j.wasman.2021.02.045

Esmaeilian, B., Wang, B., Lewis, K., Duarte, F., Ratti, C., & Behdad, S. (2018). The future of waste management in smart and sustainable cities: A review and concept paper. *Waste Manag, 81*, 177-195. doi:10.1016/j.wasman.2018.09.047

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*: CRC press.

Fatimah, Y. A., Govindan, K., Murniningsih, R., & Setiawan, A. (2020). Industry 4.0 based sustainable circular economy approach for smart waste management system to achieve sustainable development goals: A case study of Indonesia. *Journal of Cleaner Production, 269*. doi:10.1016/j.jclepro.2020.122263

Fodor, I. K. (2002). *A survey of dimension reduction techniques*. Retrieved from

Forti, V., Baldé, C., Kuehr, R., & Bel, G. (2020). *The Global E-waste Monitor 2020. Quantities, flows, and the circular economy potential*.

Garbade, D. M. J. (2018, 2018/09/13). Understanding K-means Clustering in Machine Learning. Retrieved from https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research, 11*(1), 40-56. doi:10.1016/j.apr.2019.09.009

Grossi, G., & Trunova, O. (2021). Are UN SDGs useful for capturing multiple values of smart city? *Cities, 114*. doi:10.1016/j.cities.2021.103193

Gutberlet, J. (2018). Waste in the City: Challenges and Opportunities for Urban Agglomerations. In *Urban Agglomeration*: InTech.

Gutierrez, J. M., Jensen, M., Henius, M., & Riaz, T. (2015). Smart Waste Collection System Based on Location Intelligence. *Procedia Computer Science, 61*, 120-127. doi:https://doi.org/10.1016/j.procs.2015.09.170

Hoang, A. T., Pham, V. V., & Nguyen, X. P. (2021). Integrating renewable sources into energy system for smart city as a sagacious strategy towards clean and sustainable process. *Journal of Cleaner Production, 305*. doi:10.1016/j.jclepro.2021.127161

Hossain, M. S., Al-Hamadani, S., & Rahman, M. T. (2015). E-waste: A Challenge for Sustainable Development. *J Health Pollut, 5*(9), 3-11. doi:10.5696/2156-9614-5-9.3

HRanalytics. (2017). Tutorial: How to Assess Model Accuracy. Retrieved from https://www.hranalytics101.com/how-to-assess-model-accuracy-the-basics/

Institute for Management Development, S. U. f. T. a. D. S. (2020). Smart City Index. Retrieved from https://www.imd.org/wcc/world-competitiveness-center-rankings/smart-city-index-2020/

Korea, G. o. t. R. o. (2020). *Korean New Deal*.

Kuhn, M. (2008). Building Predictive Models inRUsing thecaretPackage. *Journal of Statistical Software, 28*(5). doi:10.18637/jss.v028.i05

Kumar, A., Holuszko, M., & Espinosa, D. C. R. (2017). E-waste: An overview on generation, collection, legislation and recycling practices. *Resources, Conservation and Recycling, 122*, 32-42. doi:10.1016/j.resconrec.2017.01.018

Lee, J.-H., & Woo, J. (2020). Green New Deal Policy of South Korea: Policy Innovation for a Sustainability Transition. *Sustainability, 12*(23), 10191.

Li, X., & Xu, L. (2021). Human development associated with environmental quality in China. *PLoS One, 16*(2), e0246677. doi:10.1371/journal.pone.0246677

Lopes, M. (2017). Dimensionality Reduction — Does PCA really improve classification outcome? Retrieved from https://towardsdatascience.com/dimensionality-reduction-does-pca-really-improve-classification-outcome-6e9ba21f0a32

Marsboom, C., Vrebos, D., Staes, J., & Meire, P. (2018). Using dimension reduction PCA to identify ecosystem service bundles. *Ecological Indicators, 87*, 209-260. doi:10.1016/j.ecolind.2017.10.049

Mehta, V., Bawa, S., & Singh, J. (2020). Analytical review of clustering techniques and proximity measures. *Artificial Intelligence Review, 53*(8), 5995-6023. doi:10.1007/s10462-020-09840-7

Nazeer, K. A., & Sebastian, M. (2009). *Improving the Accuracy and Efficiency* *of the k-means Clustering Algorithm.* Paper presented at the Proceedings of the world congress on engineering.

Nistrup, P. (2019, 2019/01/29). Principal Component Analysis (PCA) 101, using R. Retrieved from https://towardsdatascience.com/principal-component-analysis-pca-101-using-r-361f4c53a9ff

OECD. Extended producer responsibility. Retrieved from https://www.oecd.org/env/tools-evaluation/extendedproducerresponsibility.htm

Ogushi, Y., & Kandlikar, M. (2007). Assessing extended producer responsibility laws in Japan. *Environ Sci Technol, 41*(13), 4502-4508. doi:10.1021/es072561x

Ottoni, M., Dias, P., & Xavier, L. H. (2020). A circular approach to the e-waste valorization through urban mining in Rio de Janeiro, Brazil. *Journal of Cleaner Production, 261*. doi:10.1016/j.jclepro.2020.120990

Programme, U. N. D. (2020). *Human Development Index* [Composite indices]. Human Development Index. Retrieved from: http://hdr.undp.org/en/content/human-development-index-hdi

Rhee, S.-W. (2016). Beneficial Use Practice of e-wastes in Republic of Korea. *Procedia Environmental Sciences, 31*, 707-714. doi:10.1016/j.proenv.2016.02.055

Serey, J., Quezada, L., Alfaro, M., Fuertes, G., Ternero, R., Gatica, G., . . . Vargas, M. (2020). Methodological Proposals for the Development of Services in a Smart City: A Literature Review. *Sustainability, 12*(24), 10249. doi:http://dx.doi.org/10.3390/su122410249

Shamsuzzoha, A., Nieminen, J., Piya, S., & Rutledge, K. (2021). Smart city for sustainable environment: A comparison of participatory strategies from Helsinki, Singapore and London. *Cities, 114*. doi:10.1016/j.cities.2021.103194

Sharma, M., Joshi, S., Kannan, D., Govindan, K., Singh, R., & Purohit, H. C. (2020). Internet of Things (IoT) adoption barriers of smart cities' waste management: An Indian context. *Journal of Cleaner Production, 270*. doi:10.1016/j.jclepro.2020.122047

Shevchenko, T., Saidani, M., Danko, Y., Golysheva, I., Chovancová, J., & Vavrek, R. (2021). Towards a Smart E-Waste System Utilizing Supply Chain Participants and Interactive Online Maps. *Recycling, 6*(1), 8. doi:10.3390/recycling6010008

Simon, D. R. (2016). Corporate Environmental Crimes and Social Inequality. *American Behavioral Scientist, 43*(4), 633-645. doi:10.1177/00027640021955469

Söder, O. (2008). k-means clustering 1. How does k-means clustering work? Retrieved from https://www.fon.hum.uva.nl/praat/manual/k-means_clustering_1__How_does_k-means_clustering_work_.html

Song, S., Duan, Y., Zhang, T., Zhang, B., Zhao, Z., Bai, X., . . . Sun, H. (2019). Serum concentrations of bisphenol A and its alternatives in elderly population living around e-waste recycling facilities in China: Associations with fasting blood glucose. *Ecotoxicol Environ Saf, 169*, 822-828. doi:10.1016/j.ecoenv.2018.11.101

Sun, X., & Yang, Z. (2008). *Generalized McNemar's test for homogeneity of the marginal distributions.* Paper presented at the SAS Global forum.

Ünlü, R., & Xanthopoulos, P. (2019). Estimating the number of clusters in a dataset via consensus clustering. *Expert systems with applications, 125*, 33-39. doi:10.1016/j.eswa.2019.01.074

Veiga, M. M. (2013). Analyzing Reverse Logistics In The Brazilian National Waste Management Policy (PNRS). In (Vol. 173, pp. 649-659). Southampton: W I T Press.

Weisstein, E. W. (2021). Eigenvalue. Retrieved from https://mathworld.wolfram.com/Eigenvalue.html

Xavier, L. H., Ottoni, M., & Lepawsky, J. (2021). Circular economy and e-waste management in the Americas: Brazilian and Canadian frameworks. *Journal of Cleaner Production, 297*.

doi:10.1016/j.jclepro.2021.126570

Zheng, B., & Agresti, A. (2000). Summarizing the predictive power of a generalized linear model. *Stat Med, 19*(13), 1771-1781. doi:10.1002/1097-0258(20000715)19:13<1771::aid-sim485>3.0.co;2-p

초 록

본 연구에서는 스마트시티의 특징과 전자폐기물 관련 지표 간 관계에 대한 데이터 분석을 진행하고자 하며, K-평균 군집의 비교사(unsupervised) 인공지능 알고리즘과 예측의 일반화된 선형 모델을 사용하여 모델 간의 비교를 통해 스마트시티 주요 특징을 파악하고자 한다. 분석 결과를 통해 전자폐기물 관리와 스마트시티 특성 간에 높은 상관관계가 관찰되었으며, 새로운 전자폐기물 재활용 처리장치 설치의 필요성을 확인하였다.

주요어: 전자 폐기물 관리, 스마트 시티, 인공지능, 클러스터링, 예측, 스마트 시티 지수, 지속 가능성입니다
학 번: 2019-23118

CERTIFICATE *of* PROOFREADING

This document certifies that the manuscript listed below was proofread for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at Express Proofreading.

Marcos Pereira de Novais

MANUSCRIPT TITLE

ARTIFICIAL INTELLIGENCE APPLIED TO SMART CITIES SELECTION FOR E-WASTE PROCESSING UNITS INSTALLATION

JULY, 21ST
2021

Express proofreading

SIGNED, *Express Proofreading*

www.EXPRESS PROOFREADING.com

Express Proofreading is a trading style of Express Communications Limited