# Artificial Intelligence implemented to recognize patterns of sustainable areas by evaluating the database of socioenvironmental safety restrictions

## Inteligência Artificial implementada para reconhecimento de padrões de áreas sustentáveis avaliando o banco de dados das restrições de segurança socioambientais

## Inteligencia Artificial se implementada para reconocer patrones de áreas sostenibles mediante la evaluación de la base de datos de restricciones de seguridad socioambientales

Received: 07/28/2021 | Reviewed: 08/02/2021 | Accept: 08/03/2021 | Published: 08/08/2021

**Julio Leite Azancort Neto**
ORCID: https://orcid.org/0000-0003-2866-5445
College Estácio Belém, Brazil
E-mail: julioazancortneto@gmail.com
**Arleson Lui Silva Gonçalves**
ORCID: https://orcid.org/0000-0001-8494-5316
College Estácio Belém, Brazil
E-mail: arlesonlsg@gmail.com
**Brennus Caio Carvalho da Cruz**
ORCID: https://orcid.org/0000-0003-2027-7894
College Estácio Belém, Brazil
E-mail: brennuscaio@gmail.com
**Larissa Luz Gomes**
ORCID: https://orcid.org/0000-0002-9086-2821
College Estácio Belém, Brazil
E-mail: larissa.gomes@estacio.br
**Denis Carlos Lima Costa**
ORCID: https://orcid.org/0000-0003-3207-6934
Federal Institute of Education, Science and Technology of Pará, Brazil
E-mail: denis.costa@ifpa.edu.br

**Abstract**

The several papers recently published, applied to sustainable development, has been considering new methodologies and techniques in identifying the main criteria, in numeric format, that are useful in formulating possible solutions to the solid waste problem. This paper presents the Mathematical and Computational Modeling Process (PM$_2$C), applied in the determination of control variables related to selection of areas destined to the construction of landfills, in order to benefit from new analyzes and values obtained by methods such as AHP (Analytical Hierarchy Process) and GIS (Geographic Information Systems). The main objective of this paper is the use of Artificial Intelligence (AI), through a Decision Tree strategy, as a selective method and optimal solutions in choosing the best area dedicated to the construction of landfills, with the creation and analysis of new values applied to scenarios defined in the paper of Andrade e Barbosa (2015). The results, expressed in analytical and graphical forms, show the individual values for each criterion and new scenarios involved in the phenomena. This paper highlights the importance of incorporating new conditions and criteria to propose a new decision-making rule, simultaneously, associating qualitative and quantitative characteristics, related to social and economic effects, applied to the environment management system. Based on these principles, it was possible to simulate new scenarios that demonstrate, with very high precision, the best values of useful criteria for decision-making in the selection of the optimal area for implementation of a landfill.

**Keywords:** Sustainable development; Environmental management; Landfill; Bio-inspired computing; Artificial intelligence; Decision tree; Decision matrix.

**Resumo**

Os diversos artigos divulgados recentemente, aplicados no desenvolvimento sustentável, têm considerado novas metodologias e técnicas na identificação dos principais critérios, em formato numérico, que são úteis na formulação de possíveis soluções para o problema dos resíduos sólidos. Este artigo apresenta o Processo de Modelagens Matemática e Computacional (PM$_2$C) aplicado na determinação das variáveis de controle relacionadas à seleção das áreas destinadas à construção de aterros sanitários, de maneira a se beneficiar das novas análises e valores obtidos pelos métodos como AHP (*Analytical Hierarchy Process*) e o GIS (*Geographic Information Systems*). O trabalho tem como objetivo primordial o uso de Inteligência Artificial (IA), mediante a estratégia de Árvore de Decisão, como método seletivo e soluções ótimas na escolha da melhor área dedicada a edificação de aterros sanitários, com a criação e análise de novos valores aplicados a cenários definidos no trabalho de Andrade e Barbosa (2015). Os resultados, expressos nas formas analítica e gráfica, exibem os valores individuais para cada critério e novos cenários envolvidos nos fenômenos. Neste artigo, denota-se a importância da incorporação de novas condições e critérios para propor novas regras de tomada de decisão, associando, simultaneamente, características qualitativas e quantitativas, relacionadas aos efeitos sociais e econômicos, aplicado ao sistema de gerenciamento ambiental. Fundamentando nesses princípios, foi possível a simulação de novos cenários que demonstram, com altíssima precisão, os melhores valores dos critérios úteis à tomada de decisão da seleção da área ótima para a implantação de um aterro sanitário.

**Palavras-chave:** Desenvolvimento sustentável; Gestão ambiental; Aterro sanitário; Computação bioinspirada; Inteligência artificial; Árvore de decisão; Matriz de critérios.

**Resumen**

Los diversos artículos publicados recientemente, aplicados al desarrollo sostenible, han considerado nuevas metodologías y técnicas en la identificación de los principales criterios, en formato numérico, que son útiles para formular posibles soluciones al problema de los residuos sólidos. Este artículo presenta el Proceso de Modelado Matemático y Computacional (PM$_2$C) aplicado en la determinación de variables de control relacionadas con la selección de áreas para la construcción de rellenos sanitarios, con el fin de beneficiarse de nuevos análisis y valores obtenidos por métodos como PJA (Proceso Analítico Jerárquico) y SIG (Sistemas de Información Geográfica). El principal objetivo del trabajo es el uso de la Inteligencia Artificial (IA), a través de la estrategia Árbol de Decisión, como método selectivo y soluciones óptimas en la elección de la mejor zona dedicada a la construcción de vertederos, con la creación y análisis de nuevos valores aplicados. a los escenarios definidos en el trabajo de Andrade y Barbosa (2015). Los resultados, expresados en forma analítica y gráfica, muestran los valores individuales para cada criterio y los nuevos escenarios involucrados en los fenómenos. Este artículo destaca la importancia de incorporar nuevas condiciones y criterios para proponer nuevas reglas de toma de decisiones, asociando simultáneamente características cualitativas y cuantitativas, relacionadas con los efectos sociales y económicos, aplicadas al sistema de gestión ambiental. A partir de estos principios, fue posible simular nuevos escenarios que demuestran, con altísima precisión, los mejores valores de los criterios útiles para la toma de decisiones en la selección de la zona óptima para la implementación de un vertedero.

**Palabras clave:** Desenvolvimiento sustentable; Gestion ambiental; Vertedero; Sistemas bioinspirados; Inteligencia artificial; Árbol de decisión; Matriz de criterios.

## 1. Introduction

In the process of maintaining the environment, efficient waste management is essential. To accomplish this, one of the main items to be defined is the location of landfills. The methodologies that determine the location must focus, fundamentally, on the prevention of risk and threats to the environment caused by short-term pollution. It is well known that the waste disposal technique is based on collection, processing, recycling, and final disposal.

Although each country has its particularities in relation to waste production, for Khorram et al (2015), in large parts of cities, waste disposal is done in a basic form of collection and deposited in landfills.

Priya et al (2019), state in their paper that, unfortunately, environmental departments have not devoted the necessary attention to the mathematization of this problem, in order to find a sui generis area for the disposal of waste.

Considering the depletion of our planet's non-renewable resources, there is an urgent need to further incorporate technology into sustainable development. It is understood that Artificial Intelligence is one of these fundamental technological modalities for this socio-environmental management.

Education is one of the fields in which the transformational potential of computing is still not well recognized. Although Bio-inspired Computing and Artificial Intelligence can be educationally attractive, there is no relevant interest in using software in teaching in order to improve the effectiveness of learning and research (Mayer, 2019).

Landfill screening is an extremely important chapter in the urban planning process, the implementation has direct impacts on the social-environmental health, ecology, and economy of the region.

Maps and geological data should be used in order to locate flaws where the structure of the crust in the region is weak. Soil maps, road maps and other environmental data sets should also be considered in the location of a safe and ecologically correct waste disposal area.

An optimal region of waste disposal should consider several characteristics in its implementation (Hayeri et al., 2019). To prevent water pollution and groundwater, threatening the ecosystem, these areas should be away from places with flooding and groundwater record.

In Brazil, despite the Nacional Solid Waste Policy demanding the end of dumps through the country, in the state of Pará, the landfill in the municipality of Marituba receives solid waste from the capital Belém, from the city of Ananindeua, from Marituba itself, from Benevides and Santa Barbara. These cities are part of the Metropolitan Region of Belém (RMB), and together they collect around 40 thousand tons per day (Brito et al, 2020).

The Marituba landfill is still in operation. Since the opening, residents of Marituba and adjacent cities have complained about the stench that invades the streets, houses and establishments. The city of Belém and its metropolitan region ate still searching for a solution for solid waste management. One of the solutions will be presented in this paper.

Decision methods aim to satisfy one or multiple objectives and are developed based on the evaluation of one or more criteria. According to Costa et al. (2020), the location of the landfill is a multi-criteria process, which considers several attributes and implies the evaluation and selection of suitable areas, based on pre-defined criteria.

For Swacha et al. (2021), there are two basic ways to integrate sustainability and education in Computing: one is by introducing new courses in the computational area whose topics across the two areas; the other is the implementation of projects and research with the theme of sustainability in classic courses, such as Computer Engineering. The second way seems more suitable. Courses whose disciplines are only indirectly linked to sustainable development issues, such as Algorithms and Data Structure or Introduction to Computer Programming, can also be seen as a solution to a recent trend of removing sustainable production proposals from the main contents of study.

This paper presents the Mathematical and Computational Modeling Process ($PM_2C$) applied in the determination of control variables related to the selection of areas for the construction of landfills. To this end, the methodology that involves technological knowledge linked to scientific knowledge was use. According to Chalmers (1999) and Crump (2002), Mathematics, Science, Engineering and Technology coexist in an evolutionary structure, proposing consistent explanations and predictions, through systematic experimental results.

## 2. Theoretical Framework

A choice between alternatives characterized a decision, which can represent different information or hypotheses about an area. The criteria serve as norms to find the best alternatives and represent possible conditions to quantify or evaluate, contributing to decision making (ABNT, 1997).

Souto (2009) describes sanitary landfills as the most viable form of final disposal of urban solid waste in Brazil, both technically and economically. However, deterring the location of landfills is a difficult and complex process, as multiple criteria must be combined to do so.

Sener et al (2011), performed the selection of suitable sites for the implementation of landfills in the catchment area of Lake Beyşehir, Turkey, using GIS and multi-criteria analysis. The survey determined eight important criteria to be considered when selecting these sites. These are distance from settlements, distance from surface water, distance from protected areas (Ecological, Scientific or Historical), Geology/Hydrogeology, land usage distance from roads, slopes and exposure.

Regarding the implementation of sanitary landfills, Portella and Ribeiro (2014) highlighted that the advantages are great, as they enable and adequate disposal of waste according to engineering and environmental control standards; high daily absorption of waste generated; they offer all the conditions for biological decomposition of organic matter contained in household or domestic waste and provide treatment for the slurry generated by the decomposition of organic matter and rainfall.

As pointed by Moreira et al. (2016), the use of sanitary landfills is the most common method of disposal of Urban Solid Waste (USW) in Brazil.

Costa et al. (2016) propose a safety region, based on Decision Tree, for the integrated dispatch of the natural gas thermoelectric power generation system. Based on the values obtained by the Decision Tree, it was possible to optimize the integrated dispatch of this system, substantially reducing the environmental impact on the ecosystem.

Kumar et al. (2017), describes that the attributes that define the area for a landfill are sets of objectives and criteria. Were the criteria being effective factors of the procedure and are operational parameters that can be scored and weighted. The evaluation of attributes provides data and information necessary to estimate alternatives to the construction of the landfill.

According to Mu et al. (2017), Decision Trees are one of the most effective and widely used techniques in many areas, such as Data Mining, Machine Learning, Image Processing and Fault Detection. According to the authors, Decision Tree has become popular not only for its high precision and need fewer parameters, but also because of its better understandability of classification rules extracted from resource-based examples, which is a very attractive property in the context of Data Mining

Wu et al. (2018), indicate in their paper, the use of Multicriteria Iterative Decision Making. According to the authors, this method is based on the Perspective Theory and considers the behavior of processes in relation to factors by which the are affected.

Pinheiro et al. (2019), characterized areas restricted to the implementation of landfills in the Pontal of Paranapanema – SP region using multi-criteria approach, applying geoprocessing tools, supervised classification, and Boolean logic. The restrictive criteria used were drainage network, water body aerodromes and Conservation Units (UC).

Sodre et al. (2020), proposed a methodology with the objective of evaluating the territory of the city of Castanhal in the state of Pará and selecting the area that best fits the current federal norms of urban waste management and sustainability. The paper highlights the fact that the city maintains an open-air dump, without any treatment, in addition to being inadequate for being close to rivers, flooding regions and houses. By using the same criteria that inappropriate the landfill region, with the aid of data from geographic information system and remote sensing, the ideal location for installing the landfill project was selected through land analysis.

Costa et al. (2021) present a Mathematical-Computational model capable of minimizing the operational costs of a multi-objective function, of thermoelectric generation, proposing the replacement of diesel oil (more pollutant) by a natural gas (80% less pollutant than diesel oil). Natural gas and electricity networks are modeled by two groups of non-linear equations and are solved by the combination of a hybrid system that applies Newton's method associated with an Artificial Intelligence strategy, called Genetic Algorithm.

## 3. Methodology

The methodology adopted in this paper is predominantly structuralist in accordance with Pereira et al (2018). The research is stablished with the investigation of a factual event. Then amplifies to the abstraction of the plan by the design of mathematical and Computational models, described by Costa et al. (2020). Computational modeling was performed with the use of RapidMiner Studio version 9.9, registered with the Educational Edition. The computer used to perform the simulations used the Windows 10 Pro Edition Version 21H1 operating system and the following specifications: CPU Ryzen 5 3600, GPU RX 580 8GB with 16GB of DDR4 memory and 480GB SDD. The scientific purpose is represented and, finally, the result of the investigation is shown, linked to a priori information about reality, idealized and correlated with social, environmental, and economic conditions and restrictions.

### 3.1 Bio-inspired Computing

Bio-inspired Computing represents a set of different studies in Computer Science, Biology and Mathematics, in addition to being a field of study of connectionism and social behavior. The Bio-inspired Computational Optimization Algorithms approach is based on the principle of biological evolution of nature to develop more robust computational techniques. In recent years, Bio-inspired Optimization Algorithms have been used in Machine Learning to address optimal solutions in solving complex problems in science and engineering.

The way that Bio-inspired Computing differs from Artificial Intelligence (AI) is the implementation of an evolutionary approach to the realization of learning, contrary to the creationist methodology of the AI. (Malladi; Shyamala, 2015).

This method can be applied from information processing, decision making to optimization algorithm. The Computational Intelligence techniques have expanded to several areas, in a way that, in the last decades, new methods and algorithms have been developed for the most different fields and applications, for example: Genetic Algorithms, Artificial Neural Networks, Evolutionary Algorithm and Fuzzy Logic.

Due to the popularization and expansion of the technology, it is expected that in the coming years, intelligent optimization algorithms will be increasingly effective in solving problems in different areas, such as: Engineering, Medicine, Space and many others.

Inspired by the working of human memory, Krestinskaya and James (2016), proposed a new Bio-inspired Algorithm to store Hierarchical Temporal Memory (HTM) resources detected in images. The proposed algorithm was tested with easy recognition using AR face database data. The simulation results showed that the proposed algorithm offers greater precision in facial recognition when compared to conventional methods.

With the beginning of the applications of Bio-inspired and Biomimetic strategies in the development of neural probes, Yang et al. (2019) presents in their article a Bio-inspired project for Neuron-Like Electronic Neural Probers (NeuE), where the main building blocks mimic the subcellular structural characteristics and mechanical properties of neurons.

### 3.2 Artificial Intelligence

Artificial Intelligence is a branch of science that seeks, through various technologies, to simulate processes in Nature aimed at solving problems. It can be found in different ways such as in technical infrastructure, processes or in a product for end users. The profound changes brought about AI in modern society are already evident in the way we live and work.

With the evolution of Computing, Artificial Intelligence has been gaining more space, as its development promoted a great advance in computational analysis, making possible the creation of technologies such as Augmented Reality, Neural Language Processing, Machine Learning, Speech Recognition, among others that allow its use in companies from different segments, to support smarter decision-making (Lu et al. 2017).

Johson et al. (2018), applied Artificial Intelligence in cardiology using a supervised learning algorithm. The author describes that, patients are starting to demand faster and more personalized care. For this, data needs a more sophisticated and efficient interpretation. The solution is Machine Learning applied to research, diagnosis and patient treatment selection, resulting in a more efficient convenient and personalized clinical practices.

Smart assistants performing a variety of tasks for their users, autonomous cars sharing learning across the entire fleet and social networks delivering content based on past behavior are clear examples of how AI is already present in our daily lives. (Szczepański, 2019).

Vaisy et al. (2020) identified seven significant AI applications for the COVID-19 pandemic. This shows that this type of technology plays and important role in detecting groups of cases and predicting where the virus will affect in the future., through the collection and analysis of previous data.

### 3.3 Decision Tree

The concept used in this paper is the Decision Tree. It is a predictive statical model of supervised learning used for data classification and prediction. For Garcia (2004, p.34), Decision Tree are a simple and effective way to represent knowledge. They are based on the divide-and-conquer approach, this means that, on the successive division of the set of examples used for training, into several subsets, until each of these subsets belongs to the same class, or until one of the classes is the majority, with no need for new divisions.

As pointed out by Crepaldi et al. (2010), the main advantage of Decision Tree is decision-making considering the most relevant attributes, as well being understandable for most people. By choosing and presenting the attributes in order of importance, Decision Tree allow users to know which factors are the most influential.

Hasan et al. (2018) used Decision Tree as a mean of predicting student performance and helping those involved in evaluating the teaching process of E-Commerce. Technologies module. In the paper, the performance of 8 Decision Trees algorithms was evaluated with a database of 22 students, the database was composed of: Academic data of each student and time spent on the Moodle online platform.

Sathiyanarayanan et al. (2019) used supervised machine learning and Decision Tree to identify breast cancer. In addition to the Decision Tree (DT) algorithm, the K-nearest Neighbors Algorithm (KNN) method was also used for precision comparison. The results reveal that despite the 97% accuracy of the KNN method, the 99% maximum accuracy of the DT is much more reliable when applied together with the Supervised Machine Learning method to predict the presence of cancer.

Freddo et al. (2019), describes the application of Decision Trees as a Data Mining method in the medical records of the Clinical School of Nutrition of the Federal University of Frontreira Sul (UFSS), Realeza campus. The technique was applied using data from 1339 medical records to identify dyslipidemia, diabetes, and hypertension, creating a tree for each. In the results, an efficiency of up to approximately 94% was obtained, showing that the rules created from a Decision Trees are valid in the identification of diseases.

De Felice et al. (2020) suggested a Decision Tree algorithm to recognize new and know clinical conditions prior to treatment needed for survival of Locally Advanced Rectal Cancer (LARC). The analysis showed that even non-specialists in the field, especially in Classification Trees, can easily interpret the tree-based Machine Learning process. Showing that Decision Tree is a way to improve decision-making in clinical practices when using larges datasets.

Ramadhan et al. (2020), performed a comparative analysis of accuracy between the K-nearest Neighbors (KNN) algorithm and Decision Tree (DT) algorithm in the detection of DDoS attacks. Using the CICIDS2017 dataset, it was possible to verify that even with KNN's 98.94% accuracy, the DT method had 99.91% of accuracy, showing which is the best method in detecting DDoS attacks.

The 2019 global pandemic of Coronavirus (COVID-19) disease has resulted in increased demand for faster and more effective testing, diagnosis,and treatment. Knowing this, Yoo et al (2020) used Deep Learning-based Decision Tree to detect COVID-19 from chest radiography images, using the definitive teste method for COVID-19, the Reverse Transcription Polymerase Chain Reaction (RT-PCR).

### 3.4 Decision Matrix

Based on the Criteria comparison matrix, defined by Andrade and Barbosa (2015), using the Analytic Hierarchy Process (AHP) method and multi-criteria analysis to determine the attributes for each analyzed variable, Costa et al. (2020) created a non-linear mathematical model, which adjusted the values of the variables *xi*, previously determined by the AHP methodology. Table 1 presents the values used in the mathematical modeling by Costa et al. (2020). This table was based on the matrix of Andrade and Barbosa (2015).

**Table 1** – Decision Matrix.

| Scenarios | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | $y_j$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | 6,00 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | 12,14 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | 6,50 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | 16,83 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | 15,50 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | 10,83 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | 11,75 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | 34,00 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | 14,50 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | 18,50 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | 13,33 |

Source: Costa et al. (2020).

According to Costa et al. (2020), “each variable *xi* assumes values related to the relevance level of the criterion calculated via AHP. The criteria are classified into geographic areas and environmental factors. These studied areas were called scenarios”.

**Table 2 –** Variables applied in MRMNL.

| Variable $x_i$ | Criterion |
|---|---|
| $x_1$ | Urban Nucleus |
| $x_2$ | Aerodrome |
| $x_3$ | Permeability |
| $x_4$ | Protected Area |
| $x_5$ | Seafront |
| $x_6$ | Streams |
| $x_7$ | Slope |
| $x_8$ | Road Network |
| $x_9$ | Exhibition/Strands |
| $x_{10}$ | Wells/Holes |
| $x_{11}$ | Cultivated Areas |

Source: Costa et al. (2020).

For the analytical and graphical presentation of the elements, Costa et al (2020), used the Nonlinear Multiple Regression Method (MRMNL), in which y*i* corresponds to the dependent variables and *xi* to the independent variables. The independent variables were named and organized to Table 2.

### 3.5 Mathematical Modeling

The Decision Matrix, defined by Andrade and Barbosa (2015), contains 11 scenarios and each is described by 11 variables. For the induction of the decision tree, this number of scenarios proved to be insufficient, so the generation of new scenarios. Therefore, each study case adopted a criterion (Arithmetic Mean, Geometric Mean and Standard Deviation) to create new scenarios.

### 3.6 Creation of New Scenarios

**Case Study A:**

Case study A used the simple Arithmetic Mean as the base criterion to produce new scenarios and split validation as the validation method.

The Arithmetic Mean can be obtained by dividing the sum of all the values of a numerical set by the total number of elements in this set, according to equation 1.

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \qquad (1)$$

where,

$\bar{x}$ → represents the value of the Arithmetic Mean

$x_i$ → the amounts involved in the sample

$f_i$ → the frequency of each $x_i$

$n$ → the quantity of $x_i$

Test 1A and 2A:

The first two tests included the insertion of 13 new scenarios (12 to 24), seeking to match the quantities for each skill, totaling 24, as shown in Table 3.

**Table 3 –** 1A and 2A Test Scenarios.

| **Scenarios** | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | **Skill** | **Arithmetic Mean** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,545 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 1,104 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,591 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 1,530 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 1,409 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,985 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 1,068 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 3,091 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 1,318 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 1,682 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 1,212 |
| **12** | 1,2 | 1/2 | 0,79 | 1/3 | 1,19 | 0,48 | 1/2 | 1/2 | 1/3 | 1/3 | 1/3 | LOW | 0,590 |
| **13** | 3,2 | 2,2 | 1,8 | 0,78 | 1,18 | 1,18 | 2,8 | 1/2 | 1,7 | 0,7 | 0,79 | LOW | 1,530 |
| **14** | 3,8 | 1,8 | 1,2 | 1,2 | 0,85 | 0,85 | 1,2 | 0,7 | 1,8 | 0,8 | 1,2 | LOW | 1,400 |
| **15** | 2,8 | 1,8 | 2,2 | 1,2 | 0,83 | 0,87 | 3,17 | 1/2 | 1,9 | 1/3 | 1,25 | LOW | 1,532 |
| **16** | 0,84 | 1/2 | 0,83 | 0,72 | 0,85 | 1/3 | 1/2 | 1/2 | 1/3 | 1/3 | 0,75 | LOW | 0,590 |
| **17** | 1,02 | 0,51 | 0,8 | 1/2 | 2/5 | 0,8 | 1/3 | 2/7 | 1/3 | 1/2 | 1/2 | HIGH | 0,545 |
| **18** | 0,8 | 0,8 | 2,2 | 1,2 | 0,8 | 0,82 | 1,2 | 1/2 | 0,9 | 0,9 | 0,7 | HIGH | 0,984 |
| **19** | 4,2 | 7,2 | 2,8 | 2,8 | 2,2 | 3,2 | 3,8 | 0,8 | 2,1 | 2,1 | 2,8 | HIGH | 3,091 |
| **20** | 2,8 | 1,8 | 2,2 | 2,2 | 0,8 | 0,8 | 3,2 | 0,7 | 0,9 | 0,9 | 2,2 | HIGH | 1,682 |
| **21** | 2,2 | 1,2 | 2,8 | 1/3 | 0,7 | 1,2 | 1,8 | 1/5 | 1/3 | 0,6 | 0,8 | MEDIUM | 1,106 |
| **22** | 1,8 | 1/3 | 3,15 | 1/2 | 0,8 | 0,8 | 1,1 | 1/2 | 1/2 | 4/9 | 1,8 | MEDIUM | 1,067 |
| **23** | 2,2 | 2,17 | 1,8 | 1/3 | 0,7 | 1,1 | 2,8 | 1/3 | 1,2 | 1,07 | 0,8 | MEDIUM | 1,319 |
| **24** | 2,8 | 0,88 | 2,2 | 1,2 | 0,8 | 1,84 | 0,7 | 1/2 | 0,87 | 1/3 | 1,2 | MEDIUM | 1,211 |

Source: Authors (2021).

The new values were determined based on already existing ones in the work of Andrade and Barbosa (2015). These values were increased or decreased up to 0.2, evenly distributed between the variables, always taking care to maintain the value of the Arithmetic Mean related to the skill of the scenario chosen as the base.

Test 3A:

Unlike Tests 1A and 2A, which have 24 scenarios, seeking to match the amount of skill between them, Test 3A now has 22 scenarios, with 11 base scenarios (Andrade and Barbosa, 2015) and 11 new ones based on these, as shown in Table 4.

**Table 4** – 3A Test Scenarios.

| Scenarios | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | Skill | Arithmetic Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,545 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 1,104 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,591 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 1,530 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 1,409 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,985 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 1,068 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 3,091 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 1,318 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 1,682 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 1,212 |
| **12** | 1,2 | 1/2 | 0,79 | 1/3 | 1,19 | 0,48 | 1/2 | 1/2 | 1/3 | 1/3 | 1/3 | LOW | 0,590 |
| **13** | 3,2 | 2,2 | 1,8 | 0,78 | 1,18 | 1,18 | 2,8 | 1/2 | 1,7 | 0,7 | 0,79 | LOW | 1,530 |
| **14** | 3,8 | 1,8 | 1,2 | 1,2 | 0,85 | 0,85 | 1,2 | 0,7 | 1,8 | 0,8 | 1,2 | LOW | 1,400 |
| **15** | 1,02 | 0,51 | 0,8 | 1/2 | 2/5 | 0,8 | 1/3 | 2/7 | 1/3 | 1/2 | 1/2 | HIGH | 0,545 |
| **16** | 0,8 | 0,8 | 2,2 | 1,2 | 0,8 | 0,82 | 1,2 | 1/2 | 0,9 | 0,9 | 0,7 | HIGH | 0,984 |
| **17** | 4,2 | 7,2 | 2,8 | 2,8 | 2,2 | 3,2 | 3,8 | 0,8 | 2,1 | 2,1 | 2,8 | HIGH | 3,091 |
| **18** | 2,8 | 1,8 | 2,2 | 2,2 | 0,8 | 0,8 | 3,2 | 0,7 | 0,9 | 0,9 | 2,2 | HIGH | 1,682 |
| **19** | 2,2 | 1,2 | 2,8 | 1/3 | 0,7 | 1,2 | 1,8 | 1/5 | 1/3 | 0,6 | 0,8 | MEDIUM | 1,106 |
| **20** | 1,8 | 1/3 | 3,15 | 1/2 | 0,8 | 0,8 | 1,1 | 1/2 | 1/2 | 4/9 | 1,8 | MEDIUM | 1,067 |
| **21** | 2,2 | 2,17 | 1,8 | 1/3 | 0,7 | 1,1 | 2,8 | 1/3 | 1,2 | 1,07 | 0,8 | MEDIUM | 1,319 |
| **22** | 2,8 | 0,88 | 2,2 | 1,2 | 0,8 | 1,84 | 0,7 | 1/2 | 0,87 | 1/3 | 1,2 | MEDIUM | 1,211 |

Source: Authors (2021).

The new scenarios have the same amount of skill as the base scenarios, with 4 HIGH, 4 MEDIUM and 3 LOW. The construction of Table 4 is based on Table 3, differing only in the LOW skill scenarios, these being 3 instead of 5. The LOW skill scenarios in Table 3 used in Table 4 are 12, 13 and 14.

Test 4A:

Just like Test 3A, Test 4A has 22 scenarios, with 11 base scenarios (Andrade and Barbosa, 2015) and 11 new ones based on these, as shown in Table 5.

**Table 5** – 4A Test Scenarios.

| **Scenarios** | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | **Skill** | **Arithmetic Mean** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,545 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 1,104 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,591 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 1,530 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 1,409 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,985 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 1,068 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 3,091 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 1,318 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 1,682 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 1,212 |
| **12** | 1,2 | 1/2 | 0,79 | 1/3 | 1,19 | 0,48 | 1/2 | 1/2 | 1/3 | 1/3 | 1/3 | LOW | 0,590 |
| **13** | 2,8 | 1,8 | 2,2 | 1,2 | 0,83 | 0,87 | 3,17 | 1/2 | 1,9 | 1/3 | 1,25 | LOW | 1,532 |
| **14** | 0,84 | 1/2 | 0,83 | 0,72 | 0,85 | 1/3 | 1/2 | 1/2 | 1/3 | 1/3 | 0,75 | LOW | 0,590 |
| **15** | 1,02 | 0,51 | 0,8 | 1/2 | 2/5 | 0,8 | 1/3 | 2/7 | 1/3 | 1/2 | 1/2 | HIGH | 0,545 |
| **16** | 0,8 | 0,8 | 2,2 | 1,2 | 0,8 | 0,82 | 1,2 | 1/2 | 0,9 | 0,9 | 0,7 | HIGH | 0,984 |
| **17** | 4,2 | 7,2 | 2,8 | 2,8 | 2,2 | 3,2 | 3,8 | 0,8 | 2,1 | 2,1 | 2,8 | HIGH | 3,091 |
| **18** | 2,8 | 1,8 | 2,2 | 2,2 | 0,8 | 0,8 | 3,2 | 0,7 | 0,9 | 0,9 | 2,2 | HIGH | 1,682 |
| **29** | 2,2 | 1,2 | 2,8 | 1/3 | 0,7 | 1,2 | 1,8 | 1/5 | 1/3 | 0,6 | 0,8 | MEDIUM | 1,106 |
| **20** | 1,8 | 1/3 | 3,15 | 1/2 | 0,8 | 0,8 | 1,1 | 1/2 | 1/2 | 4/9 | 1,8 | MEDIUM | 1,067 |
| **21** | 2,2 | 2,17 | 1,8 | 1/3 | 0,7 | 1,1 | 2,8 | 1/3 | 1,2 | 1,07 | 0,8 | MEDIUM | 1,319 |
| **22** | 2,8 | 0,88 | 2,2 | 1,2 | 0,8 | 1,84 | 0,7 | 1/2 | 0,87 | 1/3 | 1,2 | MEDIUM | 1,211 |

Source: Authors (2021).

The construction of Table 5 is based on Table 3, differing only in the LOW skill scenarios, these being 3 instead of 5. The LOW skill scenarios in Table 3 used in Table 5 are 12, 13 and 16.

Test 5A:

Just like Test 4A, Test 5A has 22 scenarios, with 11 base scenarios (Andrade and Barbosa, 2015) and 11 new ones based on these, as shown in Table 6.

**Table 6** – 5A Test Scenarios.

| **Scenarios** | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | **Skill** | **Arithmetic Mean** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,545 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 1,104 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | MEDIUM | 0,591 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 1,530 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 1,409 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,985 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 1,068 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 3,091 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 1,318 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 1,682 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 1,212 |
| **12** | 3,2 | 2,2 | 1,8 | 0,78 | 1,18 | 1,18 | 2,8 | 1/2 | 1,7 | 0,7 | 0,79 | LOW | 1,530 |
| **13** | 3,8 | 1,8 | 1,2 | 1,2 | 0,85 | 0,85 | 1,2 | 0,7 | 1,8 | 0,8 | 1,2 | LOW | 1,400 |
| **14** | 2,8 | 1,8 | 2,2 | 1,2 | 0,83 | 0,87 | 3,17 | 1/2 | 1,9 | 1/3 | 1,25 | LOW | 1,532 |
| **15** | 1,02 | 0,51 | 0,8 | 1/2 | 2/5 | 0,8 | 1/3 | 2/7 | 1/3 | 1/2 | 1/2 | HIGH | 0,545 |
| **16** | 0,8 | 0,8 | 2,2 | 1,2 | 0,8 | 0,82 | 1,2 | 1/2 | 0,9 | 0,9 | 0,7 | HIGH | 0,984 |
| **17** | 4,2 | 7,2 | 2,8 | 2,8 | 2,2 | 3,2 | 3,8 | 0,8 | 2,1 | 2,1 | 2,8 | HIGH | 3,091 |
| **18** | 2,8 | 1,8 | 2,2 | 2,2 | 0,8 | 0,8 | 3,2 | 0,7 | 0,9 | 0,9 | 2,2 | HIGH | 1,682 |
| **19** | 2,2 | 1,2 | 2,8 | 1/3 | 0,7 | 1,2 | 1,8 | 1/5 | 1/3 | 0,6 | 0,8 | MEDIUM | 1,106 |
| **20** | 1,8 | 1/3 | 3,15 | 1/2 | 0,8 | 0,8 | 1,1 | 1/2 | 1/2 | 4/9 | 1,8 | MEDIUM | 1,067 |
| **21** | 2,2 | 2,17 | 1,8 | 1/3 | 0,7 | 1,1 | 2,8 | 1/3 | 1,2 | 1,07 | 0,8 | MEDIUM | 1,319 |
| **22** | 2,8 | 0,88 | 2,2 | 1,2 | 0,8 | 1,84 | 0,7 | 1/2 | 0,87 | 1/3 | 1,2 | MEDIUM | 1,211 |

Source: Authors (2021).

The construction of Table 6 is based on Table 3, differing only in the LOW skill scenarios, these being 3 instead of 5. The LOW skill scenarios in Table 3 used in Table 6 are 13, 14 and 15.

**Case Study B:**

Case study B adopted the use of Geometric Mean as a base criterion to produce new scenarios. The Geometric Mean was calculated according to equation 2.

$$\bar{g} = \sqrt[n]{\prod_{i=i}^{n} {x_i}^{f_i}} \tag{2}$$

where,

$\bar{g}$ → represents the value of the Geometric Mean.

Test 1B:

The first test included the insertion of 13 new scenarios (12 to 24), totaling 24, as shown in Table 7.

**Table 7** - 1B Test Scenarios.

| Scenarios | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | Skill | Geometric Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,47677 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 0,81623 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,54080 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 1,25345 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 1,20808 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,90495 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 0,76892 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 2,76624 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 1,10503 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 1,47519 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 1 |
| **12** | 1,2 | 0,53 | 1,2 | 0,3 | 1,2 | 0,3 | 3/4 | 1/8 | 0,7 | 0,3 | 0,7 | LOW | 0,54226 |
| **13** | 3,2 | 2 | 2,2 | 0,8 | 1,2 | 0,8 | 3,2 | 0,53 | 2,2 | 0,3 | 1 | LOW | 1,25515 |
| **14** | 3,8 | 2,2 | 1 | 1,2 | 0,8 | 1 | 0,8 | 0,7 | 1,8 | 1,2 | 0,8 | LOW | 1,20483 |
| **15** | 3,8 | 2,2 | 2,2 | 0,8 | 1 | 1,2 | 3/4 | 0,7 | 2 | 0,5 | 0,8 | LOW | 1,19684 |
| **16** | 1,2 | 1/3 | 0,8 | 0,5 | 0,8 | 0,3 | 1/3 | 1/8 | 0,7 | 1,2 | 1 | LOW | 0,54815 |
| **17** | 1 | 0,3 | 1,2 | 0,5 | 0,45 | 0,8 | 0,7 | 1/8 | 0,7 | 1/8 | 0,53 | HIGH | 0,47568 |
| **18** | 3,8 | 7,2 | 2,8 | 3,2 | 2,2 | 3,2 | 4,2 | 0,8 | 1,8 | 2,2 | 3,2 | HIGH | 2,76936 |
| **19** | 3,2 | 2 | 2,2 | 1,8 | 0,8 | 1 | 2,8 | 0,7 | 0,8 | 1,2 | 1,8 | HIGH | 1,46885 |
| **20** | 0,8 | 1 | 1,8 | 0,8 | 1 | 0,8 | 1,2 | 0,53 | 0,8 | 1,2 | 0,7 | HIGH | 0,91930 |
| **21** | 2,8 | 0,8 | 2,2 | 0,8 | 0,8 | 2,2 | 0,3 | 0,53 | 1,2 | 0,7 | 1,2 | MEDIUM | 1,01027 |
| **22** | 1,8 | 0,3 | 3,2 | 0,53 | 1 | 0,8 | 1,2 | 0,45 | 0,53 | 1/8 | 2,2 | MEDIUM | 0,77695 |
| **23** | 2,2 | 1,2 | 2,8 | 0,7 | 0,3 | 0,8 | 1,8 | 1/7 | 0,5 | 0,7 | 1 | MEDIUM | 0,81937 |
| **24** | 1,8 | 2,2 | 1,8 | 0,3 | 0,7 | 0,8 | 2,8 | 0,7 | 0,8 | 1 | 1,2 | MEDIUM | 1,07664 |

Source: Authors (2021).

Of the new data, 11 were created by adding or subtracting 0.2 of each variable from the scenarios in Table 7, maintaining the Geometric Mean value. The remaining 2 scenarios were inserted to balance the number of cases with LOW aptitude and for this, 11 values were chosen, used in the 22 previous scenarios, which resulted in one of the Geometric Averages of LOW aptitude.

Test 2B:

Test 2B had a similar table to Test 1B, containing 24 scenarios, as shown in Table 8.

**Table 8** - 2B Test Scenarios.

| Scenarios | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | Skill | Geometric Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,47677 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 0,81623 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,54080 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 1,25345 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 1,20808 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,90495 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 0,76892 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 2,76624 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 1,10503 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 1,47519 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 1 |
| **12** | 4 | 1,4 | 2,4 | 0,6 | 1 | 0,6 | 3,4 | 1/3 | 2,4 | 1 | 0,6 | LOW | 1,20665 |
| **13** | 0,5 | 2,4 | 1,6 | 0,9 | 1,6 | 0,6 | 1,6 | 0,9 | 3 | 1,6 | 1 | LOW | 1,24827 |
| **14** | 1,4 | 1/3 | 1,4 | 0,9 | 1,4 | 0,9 | 8/15 | 1/3 | 0,1 | 0,9 | 0,1 | LOW | 0,54198 |
| **15** | 3,4 | 2 | 2,4 | 0,6 | 1 | 1 | 3 | 1/3 | 2,4 | 0,9 | 0,6 | LOW | 1,25984 |
| **16** | 4,4 | 2,4 | 0,6 | 1,4 | 0,6 | 0,6 | 1,4 | 0,9 | 2,4 | 0,6 | 1,4 | LOW | 1,20961 |
| **17** | 0,6 | 0,1 | 0,6 | 3/4 | 2/3 | 0,6 | 0,9 | 2/3 | 0,1 | 3/4 | 3/4 | HIGH | 0,48168 |
| **18** | 3,6 | 6,6 | 2,6 | 2,6 | 2,4 | 2,6 | 3,6 | 1,4 | 2,4 | 1,6 | 3,4 | HIGH | 2,74227 |
| **19** | 2,6 | 2,4 | 2,4 | 1,6 | 0,6 | 1,4 | 2,6 | 0,9 | 0,6 | 1,4 | 1,6 | HIGH | 1,45802 |
| **20** | 1,4 | 1 | 2,4 | 1,4 | 0,6 | 1 | 1,4 | 3/4 | 1,4 | 1 | 0,5 | HIGH | 1,06640 |
| **21** | 2,6 | 1,4 | 1,6 | 1,4 | 0,6 | 2 | 0,9 | 3/4 | 1 | 0,1 | 1,4 | MEDIUM | 0,99087 |
| **22** | 2 | 0,1 | 2,6 | 1/3 | 1,4 | 0,6 | 1,4 | 2/3 | 3/4 | 1/3 | 1,6 | MEDIUM | 0,76405 |
| **23** | 2 | 1 | 2,6 | 0,5 | 0,9 | 1,4 | 2,4 | 1/7 | 0,1 | 0,9 | 1 | MEDIUM | 0,81192 |
| **24** | 1,6 | 2,4 | 2,4 | 0,9 | 0,9 | 1 | 3,4 | 0,1 | 1,4 | 0,6 | 1,4 | MEDIUM | 1,10452 |

Source: Authors (2021).

Of the new data, 11 were from Table 1 and 13 were new scenarios, created to enrich the database. The only difference is in the constant chosen to add or subtract the variables, which went from 0.2 to 0.4.

**Case Study C:**

According to Martins (2013), "Standard Deviation of a sample (or collection) of data, of a quantitative type, is a measure of data dispersion relative to the mean, which is obtained by taking the square root of the sample variance".

The disseminations of observations that make up a sample can be characterized by the deviations of each observation in relation to the mean $(x_i - x)$, which can take positive or negative values, and the sum of the deviations of each observation in relation to the sample mean is zero.

Based on this information, case study C used equation 3 to calculate the standard deviation and use it as a base criterion for producing new scenarios. Cross Validation and Split Validation were adopted as the validation methods. The calculation of Standard Deviation is exemplified in the equation below.

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i - 1}} \tag{3}$$

where,

$s$ → represents the Stardard Deviation value of the data involved.

Test 1C:

The first test of the case study C has the addition of 22 new scenarios (12 to 33), based on the Standard Deviation values from the original table of 11 values defined by Andrade and Barbosa (2015), as shown in Table 9.

**Table 9** - 1C Test Scenario.

| **Scenarios** | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | **Skill** | **Standard Deviation** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,306 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 0,873 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,272 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 0,927 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 0,970 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,411 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 0,900 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 1,578 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 0,815 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 0,845 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 0,803 |
| **12** | 1,2 | 0,7 | 1,2 | 0,533 | 0,45 | 1,2 | 0,7 | 0,45 | 0,7 | 0,533 | 0,533 | HIGH | 0,306 |
| **13** | 2,2 | 1,2 | 3,2 | 0,7 | 0,7 | 1,2 | 2,2 | 0,342 | 0,7 | 0,7 | 1,2 | MEDIUM | 0,873 |
| **14** | 1,2 | 0,533 | 1,2 | 0,7 | 1,2 | 0,7 | 0,533 | 0,533 | 0,7 | 0,7 | 0,7 | LOW | 0,273 |
| **15** | 3,2 | 2,2 | 2,2 | 1,2 | 1,2 | 1,2 | 3,2 | 0,533 | 2,2 | 0,7 | 1,2 | LOW | 0,927 |
| **16** | 4,2 | 2,2 | 1,2 | 1,2 | 1,2 | 1,2 | 1,2 | 0,7 | 2,2 | 1,2 | 1,2 | LOW | 0,970 |
| **17** | 1,2 | 1,2 | 2,2 | 1,2 | 1,2 | 1,2 | 1,2 | 0,533 | 1,2 | 1,2 | 0,7 | HIGH | 0,411 |
| **18** | 2,2 | 0,7 | 3,2 | 0,533 | 1,2 | 1,2 | 1,2 | 0,45 | 0,533 | 0,533 | 2,2 | MEDIUM | 0,900 |
| **19** | 4,2 | 7,2 | 3,2 | 3,2 | 2,2 | 3,2 | 4,2 | 1,2 | 2,2 | 2,2 | 3,2 | HIGH | 1,578 |
| **20** | 2,2 | 2,2 | 2,2 | 0,7 | 0,7 | 1,2 | 3,115 | 0,7 | 1,2 | 1,2 | 2,2 | MEDIUM | 0,815 |
| **21** | 3,2 | 2,2 | 2,2 | 2,2 | 1,2 | 1,2 | 3,2 | 0,7 | 1,2 | 1,2 | 2,2 | HIGH | 0,845 |
| **22** | 3,2 | 1,2 | 2,2 | 1,2 | 1,2 | 2,2 | 0,7 | 0,533 | 1,2 | 0,7 | 1,2 | MEDIUM | 0,803 |
| **23** | 1,3 | 0,8 | 1,3 | 0,633 | 0,55 | 1,3 | 0,8 | 0,55 | 0,8 | 0,633 | 0,633 | HIGH | 0,306 |
| **24** | 2,3 | 1,3 | 3,3 | 0,8 | 0,8 | 1,3 | 2,3 | 0,442 | 0,8 | 0,8 | 1,3 | MEDIUM | 0,873 |
| **25** | 1,3 | 0,65 | 1,3 | 0,8 | 1,3 | 0,8 | 0,633 | 0,633 | 0,8 | 0,8 | 0,8 | LOW | 0,271 |
| **26** | 3,3 | 2,2 | 2,3 | 1,3 | 1,3 | 1,3 | 3,3 | 0,633 | 2,3 | 0,8 | 1,3 | LOW | 0,923 |
| **27** | 4,3 | 2,3 | 1,3 | 1,3 | 1,3 | 1,3 | 1,3 | 0,8 | 2,3 | 1,3 | 1,3 | LOW | 0,970 |
| **28** | 1,3 | 1,3 | 2,3 | 1,3 | 1,3 | 1,3 | 1,3 | 0,633 | 1,3 | 1,3 | 0,8 | HIGH | 0,411 |
| **29** | 2,3 | 0,8 | 3,3 | 0,633 | 1,3 | 1,3 | 1,3 | 0,55 | 0,633 | 0,633 | 2,3 | MEDIUM | 0,900 |
| **30** | 4,3 | 7,3 | 3,3 | 3,3 | 2,3 | 3,3 | 4,3 | 1,3 | 2,3 | 2,3 | 3,3 | HIGH | 1,578 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **31** | 2,3 | 2,3 | 2,3 | 0,8 | 0,8 | 1,3 | 3,3 | 0,8 | 1,3 | 1,3 | 1,3 | MEDIUM | 0,815 |
| **32** | 3,3 | 2,3 | 2,3 | 2,3 | 1,3 | 1,3 | 3,3 | 0,8 | 1,3 | 1,3 | 2,3 | HIGH | 0,845 |
| **33** | 3,3 | 1,3 | 2,3 | 1,3 | 1,3 | 2,3 | 0,8 | 0,633 | 1,3 | 0,8 | 1,3 | MEDIUM | 0,803 |

Source: Authors (2021).

Of the 22 new values, 11 were created with the addition of 0.2 and the other 11 were created by adding 0.3 to the original values, always taking care to maintain the Standard Deviation value equal to or near the Table 1 values.

Test 2C:

The 2C test had 33 new scenarios (12 to 44), as shown in Table 10.

**Table 10** – 2C Test Scenarios.

| Scenarios | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | Skill | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,306 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 0,873 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,272 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 0,927 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 0,970 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,411 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 0,900 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 1,578 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 0,815 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 0,845 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 0,803 |
| **12** | 1,02 | 0,52 | 1,02 | 0,353 | 0,27 | 1,02 | 0,52 | 0,27 | 0,52 | 0,353 | 0,353 | HIGH | 0,306 |
| **13** | 2,02 | 1,02 | 3,02 | 0,52 | 0,52 | 1,02 | 2,02 | 0,1629 | 0,52 | 0,52 | 1,02 | MEDIUM | 0,873 |
| **14** | 1,02 | 0,353 | 1,02 | 0,52 | 1,02 | 0,52 | 0,3533 | 0,353 | 0,52 | 0,52 | 0,52 | LOW | 0,272 |
| **15** | 3,02 | 2,02 | 2,02 | 1,02 | 1,02 | 1,02 | 3,02 | 0,3533 | 2,02 | 0,52 | 1,02 | LOW | 0,927 |
| **16** | 4,02 | 2,02 | 1,02 | 1,02 | 1,02 | 1,02 | 1,02 | 0,52 | 2,02 | 1,02 | 1,02 | LOW | 0,970 |
| **17** | 1,02 | 1,02 | 2,02 | 1,02 | 1,02 | 1,02 | 1,02 | 0,3533 | 1,02 | 1,02 | 0,52 | HIGH | 0,411 |
| **18** | 2,02 | 0,52 | 3,02 | 0,353 | 1,02 | 1,02 | 1,02 | 0,27 | 0,3533 | 0,3533 | 2,02 | MEDIUM | 0,900 |
| **19** | 4,02 | 7,02 | 3,02 | 3,02 | 2,02 | 3,02 | 4,02 | 1,02 | 2,02 | 2,02 | 3,02 | HIGH | 1,578 |
| **20** | 2,02 | 2,02 | 2,02 | 0,52 | 0,52 | 1,02 | 3,02 | 0,52 | 1,02 | 1,02 | 1,02 | MEDIUM | 0,815 |
| **21** | 3,02 | 2,02 | 2,02 | 2,02 | 1,02 | 1,02 | 3,02 | 0,52 | 1,02 | 1,02 | 2,02 | HIGH | 0,845 |
| **22** | 3,02 | 1,02 | 2,02 | 1,02 | 1,02 | 2,02 | 0,52 | 0,3533 | 1,02 | 0,52 | 1,02 | MEDIUM | 0,803 |
| **23** | 1,01 | 0,51 | 1,01 | 0,3433 | 0,26 | 1,01 | 0,51 | 0,26 | 0,51 | 0,3433 | 0,3433 | HIGH | 0,306 |
| **24** | 2,01 | 1,01 | 3,01 | 0,51 | 0,51 | 1,01 | 2,01 | 0,1529 | 0,51 | 0,51 | 1,01 | MEDIUM | 0,873 |
| **25** | 1,01 | 0,3433 | 1,01 | 0,51 | 1,01 | 0,51 | 0,3433 | 0,3433 | 0,51 | 0,51 | 0,51 | LOW | 0,272 |
| **26** | 3,01 | 2,01 | 2,01 | 1,01 | 1,01 | 1,01 | 3,01 | 0,3433 | 2,01 | 0,51 | 1,01 | LOW | 0,927 |
| **27** | 4,01 | 2,01 | 1,01 | 1,01 | 1,01 | 1,01 | 1,01 | 0,51 | 2,01 | 1,01 | 1,01 | LOW | 0,970 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **28** | 1,01 | 1,01 | 2,01 | 1,01 | 1,01 | 1,01 | 1,01 | 0,3433 | 1,01 | 1,01 | 0,51 | HIGH | 0,411 |
| **29** | 2,01 | 0,51 | 3,01 | 0,3433 | 1,01 | 1,01 | 1,01 | 0,26 | 0,3433 | 0,3433 | 2,01 | MEDIUM | 0,900 |
| **30** | 4,01 | 7,01 | 3,01 | 3,01 | 2,01 | 3,01 | 4,01 | 1,01 | 2,01 | 2,01 | 3,01 | HIGH | 1,578 |
| **31** | 2,01 | 2,01 | 2,01 | 0,51 | 0,51 | 1,01 | 3,01 | 0,51 | 1,01 | 1,01 | 1,01 | MEDIUM | 0,815 |
| **32** | 3,01 | 2,01 | 2,01 | 2,01 | 1,01 | 1,01 | 3,01 | 0,51 | 1,01 | 1,01 | 2,01 | HIGH | 0,845 |
| **33** | 3,01 | 1,01 | 2,01 | 1,01 | 1,01 | 2,01 | 0,51 | 0,3433 | 1,01 | 0,51 | 1,01 | MEDIUM | 0,803 |
| **34** | 1,015 | 0,515 | 1,015 | 0,3483 | 0,265 | 1,015 | 0,515 | 0,265 | 0,515 | 0,3483 | 0,3483 | HIGH | 0,306 |
| **35** | 2,015 | 1,015 | 3,015 | 0,515 | 0,515 | 1,015 | 2,015 | 0,1579 | 0,515 | 0,515 | 1,015 | MEDIUM | 0,873 |
| **36** | 1,015 | 0,3483 | 1,015 | 0,515 | 1,015 | 0,515 | 0,3483 | 0,3483 | 0,515 | 0,515 | 0,515 | LOW | 0,272 |
| **37** | 3,015 | 2,015 | 2,015 | 1,015 | 1,015 | 1,015 | 3,015 | 0,3483 | 2,015 | 0,515 | 1,015 | LOW | 0,927 |
| **38** | 4,015 | 2,015 | 1,015 | 1,015 | 1,015 | 1,015 | 1,015 | 0,515 | 2,015 | 1,015 | 1,015 | LOW | 0,970 |
| **39** | 1,015 | 1,015 | 2,015 | 1,015 | 1,015 | 1,015 | 1,015 | 0,3483 | 1,015 | 1,015 | 0,515 | HIGH | 0,411 |
| **40** | 2,015 | 0,515 | 3,015 | 0,3483 | 1,015 | 1,015 | 1,015 | 0,265 | 0,3483 | 0,3483 | 2,015 | MEDIUM | 0,900 |
| **41** | 4,015 | 7,015 | 3,015 | 3,015 | 2,015 | 3,015 | 4,015 | 1,015 | 2,015 | 2,015 | 3,015 | HIGH | 1,578 |
| **42** | 2,015 | 2,015 | 2,015 | 0,515 | 0,515 | 1,015 | 3,015 | 0,515 | 1,015 | 1,015 | 1,015 | MEDIUM | 0,815 |
| **43** | 3,015 | 2,015 | 2,015 | 2,015 | 1,015 | 1,015 | 3,015 | 0,515 | 1,015 | 1,015 | 2,015 | HIGH | 0,845 |
| **44** | 3,015 | 1,015 | 2,015 | 1,015 | 1,015 | 2,015 | 0,515 | 0,3483 | 1,015 | 0,515 | 1,015 | MEDIUM | 0,803 |

Source: Authors (2021).

New scenarios were created to verify the impact of the quantity and proximity of data on the accuracy of the Decision Tree. The new scenarios created were: 11 values ranging from 0.01, 11 values ranging from 0.02 and 11 values ranging from 0.015.

Test 3C:

In the third test of case study C, only 11 new values were created, varying at most by 0.1 (adding or subtracting) from the original table, as shown in Table 11.

**Table 11** - 3C Test Scenarios.

| Scenarios | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ | $x_{11}$ | Skill | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 1 | 0,5 | 1 | 1/3 | 0,25 | 1 | 0,5 | 0,25 | 0,5 | 1/3 | 1/3 | HIGH | 0,306 |
| **2** | 2 | 1 | 3 | 0,5 | 0,5 | 1 | 2 | 1/7 | 0,5 | 0,5 | 1 | MEDIUM | 0,873 |
| **3** | 1 | 1/3 | 1 | 0,5 | 1 | 0,5 | 1/3 | 1/3 | 0,5 | 0,5 | 0,5 | LOW | 0,272 |
| **4** | 3 | 2 | 2 | 1 | 1 | 1 | 3 | 1/3 | 2 | 0,5 | 1 | LOW | 0,927 |
| **5** | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 0,5 | 2 | 1 | 1 | LOW | 0,970 |
| **6** | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1/3 | 1 | 1 | 0,5 | HIGH | 0,411 |
| **7** | 2 | 0,5 | 3 | 1/3 | 1 | 1 | 1 | 0,25 | 1/3 | 1/3 | 2 | MEDIUM | 0,900 |
| **8** | 4 | 7 | 3 | 3 | 2 | 3 | 4 | 1 | 2 | 2 | 3 | HIGH | 1,578 |
| **9** | 2 | 2 | 2 | 0,5 | 0,5 | 1 | 3 | 0,5 | 1 | 1 | 1 | MEDIUM | 0,815 |
| **10** | 3 | 2 | 2 | 2 | 1 | 1 | 3 | 0,5 | 1 | 1 | 2 | HIGH | 0,845 |
| **11** | 3 | 1 | 2 | 1 | 1 | 2 | 0,5 | 1/3 | 1 | 0,5 | 1 | MEDIUM | 0,803 |
| **12** | 3,03 | 2,1 | 1,9 | 0,73 | 1,1 | 1,1 | 2,9 | 4/9 | 1,9 | 0,6 | 0,98 | LOW | 0,898 |
| **13** | 3,9 | 1,9 | 1,1 | 1,1 | 0,9 | 0,9 | 1,1 | 0,6 | 1,9 | 0,9 | 1,1 | LOW | 0,922 |
| **14** | 2,9 | 1,98 | 2,1 | 1,1 | 0,9 | 0,9 | 3,09 | 4/9 | 1,9 | 0,38 | 1,15 | LOW | 0,927 |
| **15** | 3,9 | 6,9 | 3,05 | 3,03 | 2,9 | 2,05 | 3,95 | 1 | 2,06 | 1,99 | 3,05 | HIGH | 1,539 |
| **16** | 4 | 7,02 | 2,95 | 3,96 | 3,05 | 1,97 | 4,04 | 1,06 | 1,91 | 2,07 | 3,09 | HIGH | 1,601 |
| **17** | 4,1 | 7,1 | 2,9 | 2,9 | 1,9 | 3,1 | 4,1 | 0,9 | 1,98 | 2,05 | 2,9 | HIGH | 1,636 |
| **18** | 4,05 | 7 | 3,09 | 2,95 | 1,98 | 2,1 | 3,9 | 1,02 | 2,1 | 1,92 | 2,96 | HIGH | 1,602 |
| **19** | 2,1 | 1,1 | 2,9 | 0,45 | 0,6 | 1,1 | 1,9 | 0,16 | 0,45 | 0,55 | 0,9 | MEDIUM | 0,849 |
| **20** | 3,1 | 0,9 | 1,95 | 0,9 | 0,9 | 2,1 | 0,6 | 1/3 | 0,9 | 0,4 | 0,9 | MEDIUM | 0,848 |
| **21** | 2,1 | 2,07 | 1,9 | 0,42 | 0,57 | 0,9 | 2,9 | 0,42 | 1,1 | 1,09 | 0,9 | MEDIUM | 0,815 |
| **22** | 3,1 | 1,1 | 2,1 | 0,9 | 0,9 | 1,9 | 0,6 | 4/9 | 0,9 | 0,42 | 1 | MEDIUM | 0,822 |

Source: Authors (2021).

Unlike the 2 previous scenarios, this one has approximate values defined in accordance with Table 1, so that the Standard Deviation values of the LOW scenarios were between 0.898 and 0.927, the MEDIUM ones between 0.815 and 0.849 and the HIGH ones between 1.539 and 1.602. This approximation was made respecting the values in the table by Andrade and Barbosa (2015), however, making the ranges of values clear and facilitating the identification and prediction of each aptitude in the Decision Tree training.

### 3.6 Computational Modeling

Computational Modeling and Simulation is a fundamental technique for the study. With the purpose of delimiting ideal values and variables that configure a phenomenon. Paula et al. (2020), used the software, FLUENT 14, in Computational Simulations for soy particle flux form simulated analyzes. Making it possible to identify parameters to improve structures used in storage units of agricultural products.

GEOSLOPE, was the software used by Magalhães et al. (2020) in the investigation of landfill stability determined through a global analysis, using the Bishop method.

Costa et al. (2020), used MATLAB Programming Language (MATrix LABoratory), to solve the system of linear equations, adjust the data of a surface and find a function that would model the variables involved and optimize the solution of the Criteria Matix. The choice of language was made because it is a high-performance language, duly endorsed scientifically validated and presented all the necessary tools to carry out its research.

The software, RapidMiner, was chosen to carry out the research proposed in this paper, as it provides a wide range of statistical evaluation methods, such as correlation analysis for regression, classification, and clustering procedures, as well as parameter optimization. Such methods can be used in different applications and data types, such as text, images, audio, and time series analysis. The analyzes can be fully automated and the result viewed in different ways.

### 3.6.1 RapidMiner

RapidMiner is one of the world’s most widely used open-source software for data mining solutions. The project was born at the University of Dortmund in 2001 and its development has continued by Rapid-I GmbH since 2007. With its academic context, RapidMiner continues to address not only business customers, but alto in university environments and researchers of the most diverse areas.

Its use includes professionals in the fields of Computer Science, Statistics, Mathematics and other who are interested in Data Mining techniques, Machine Learning, and statistical methods. RapidMiner facilitates the implementation of new analysis methodologies and Data Mining approaches for different areas of Mathematics and Statistics, without the requirement of programming knowledge.

### 3.6.2 Decision Tree Implementation

For the Decision Tree construction, five operators were used, according to the following specifications:

a) *Read Excel* - This operator can be used to load data from Microsoft Excel spreadsheets.

b) *Split Validation* - This operator performs a simple validation i.e., randomly splits up the ExampleSet into a training set and test set and evaluates the model. This operator performs a split validation in order to estimate the performance of a learning operator (usually on unseen data sets). It is mainly used to estimate how accurately a model (learnt by a particular learning operator) will perform in practice.

c) *Cross Validation* - It is mainly used to estimate how accurately a model (learned by a particular learning Operator) will perform in practice. It has two subprocesses: a Training subprocess and a Testing subprocess. The input ExampleSet is partitioned into k subsets of equal size. Of the k subsets, a single subset is retained as the test data set (i.e., input of the Testing subprocess). The remaining k - 1 subsets are used as training data set (i.e., input of the Training subprocess). The cross-validation process is then repeated k times, with each of the k subsets used exactly once as the test data.

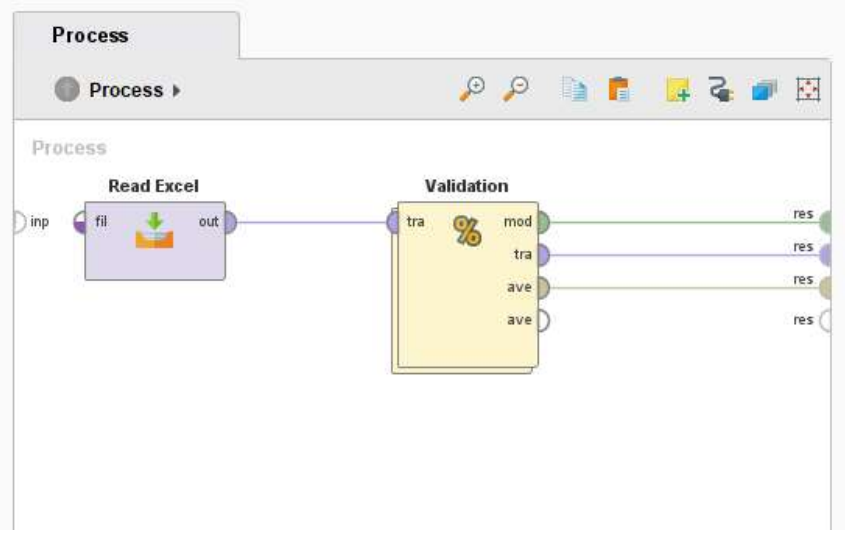
d) *Decision Tree* - This Operator generates a decision tree model, which can be used for classification and regression.

e) *Apply Model* - A model is first trained on an ExampleSet by another Operator, which is often a learning algorithm. Afterwards, this model can be applied on another ExampleSet. Usually, the goal is to get a prediction on unseen data or to transform data by applying a preprocessing model.

f) *Performance* - This operator is used for statistical performance evaluation of classification tasks. This operator delivers a list of performance criteria values of the classification task.

The following figures represents the programming of the operators mentioned above, as well as their connections. These connections are in accordance with the methodology used for this DT, specifically.

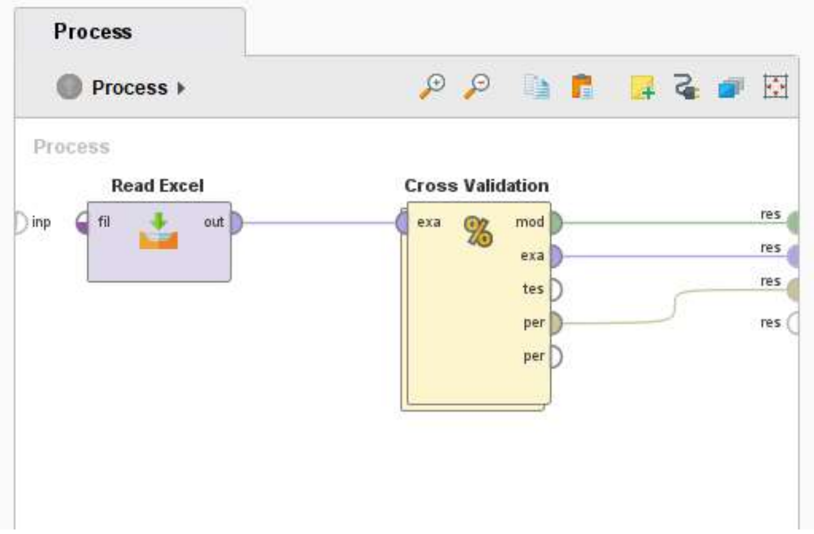
**Figure 1 -** Connection Structure Using *Split Validation.*



Source: Authors (2021).

Figure 1 represents the programming of the Read Excel and Split Validation operators as well as their connections. These connections are in accordance with the methodology used for this DT, specifically.

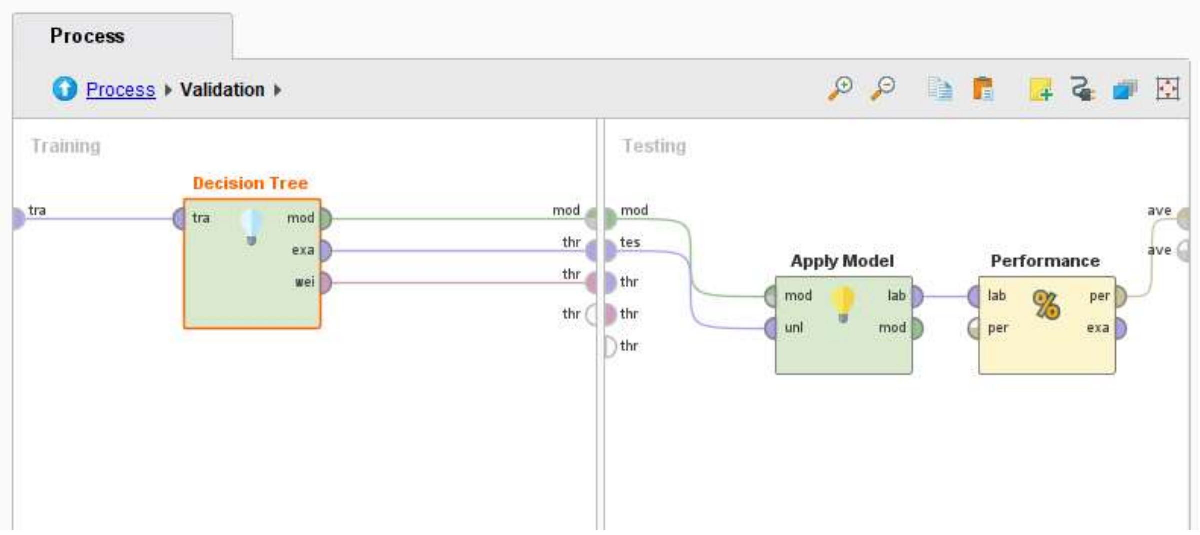
**Figure 2 -** Connection Structure Using *Cross Validation.*



Source: Authors (2021).

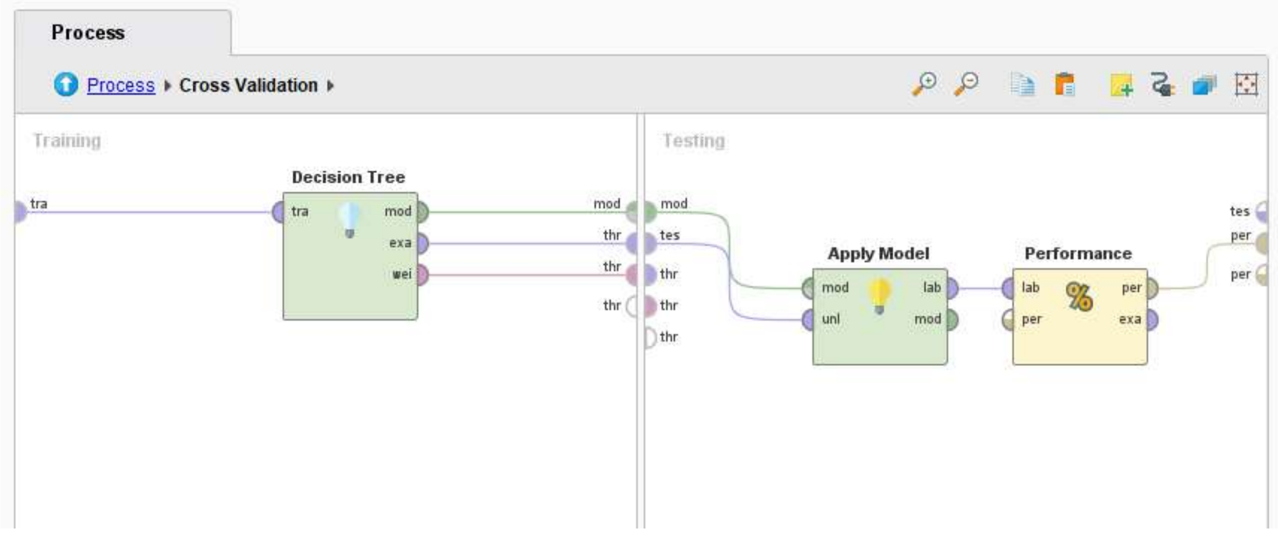
Figure 2 represents the programming of the Read Excel and Cross Validation operators as well as their connections. These connections are in accordance with the methodology used for this DT, specifically.

**Figure 3 -** Subroutines *Split Validation.*



Source: Authors (2021).

**Figure 4 -** Subroutines *Cross Validation*.



Source: Authors (2021).

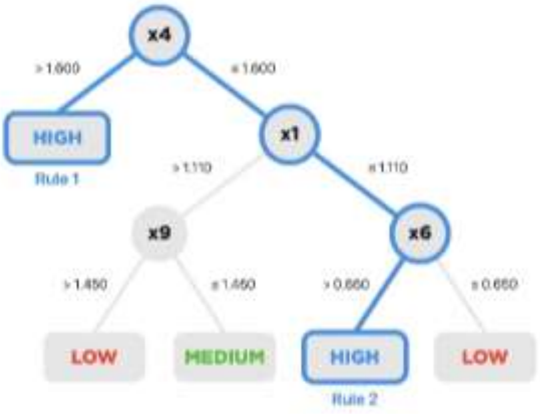
Figures 3 and 4 represent the programming of Decision Tree, Apply Model and Performance operators as well as their connections according to the validation method used. These operators are subroutines of the main validation operator and are contained within it. The connections are in accordance with the methodology used for this DT, specifically.

## 4. Results and Discussion

**Case 1A:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 3, we obtained the DT presented in Figure 5 and Table 12.

**Figure 5 -** DT referring to Case 1A.



Source: Authors (2021).

According to Figure 5, Rules 1 and 2 define the control variables x4, x1 and x6, in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (4) and (5).

$$x4 > 1.600 \tag{4}$$

$$x4 \leq 1.600 \ \rightarrow x1 \leq 1.110 \ \rightarrow x6 > 0.650 \tag{5}$$

where,

$x_1$ → refers to the distance of Urban Nucleus;

$x_4$ → refers to the distance of Protected Areas;

$x_6$ → refers to the distance of Streams;

**Table 12 -** Performance Vector 1A.

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 2 | 0 | 1 | 66,67% |
| **pred. MEDIUM** | 0 | 2 | 0 | 100,00% |
| **pred. LOW** | 0 | 0 | 1 | 100,00% |
| **class recall** | 100,00% | 100,00% | 50,00% | |
| **accuracy** | **83,33%** | | | |

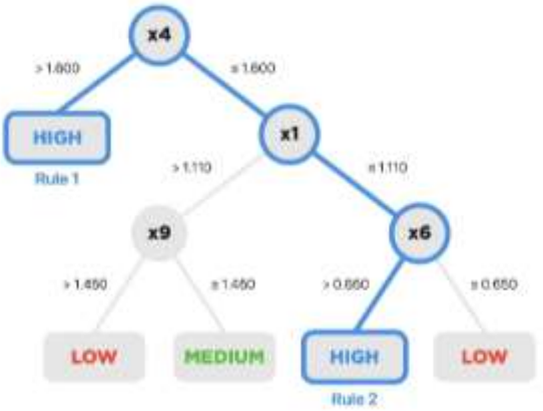
Source: Authors (2021).

The Performance Vector shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 12, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 66,67%, 100% and 100% respectively, and a class recall of 100%, 100% and 50% respectively, resulting in an accuracy of 83,33%.

**Case 2A:**

In order to improve the accuracy of the results obtained through Table 3, the parameters of the operators used in this case are in their default format, with the exception of the Split Ratio, which was changed from its default 0.7 to 0.6. After executing the process, with the data presented in Table 3, we obtained the DT presented in Figure 6 and Table 13.

**Figure 6 -** DT referring to Case 2A.



Source: Authors (2021).

According to Figure 6, the DT has the same rules presented by the DT in Figure 5, as well as their expressions.

**Table 13 -** Performance Vector 2A.

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 3 | 0 | 1 | 75,00% |
| **pred. MEDIUM** | 0 | 3 | 0 | 100,00% |
| **pred. LOW** | 0 | 0 | 2 | 100,00% |
| **class recall** | 100,00% | 100,00% | 66,67% | |
| **accuracy** | | **88,89%** | | |

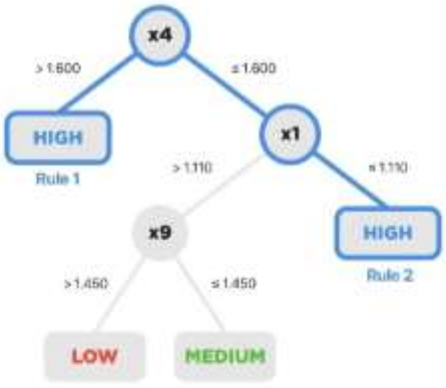
Source: Authors (2021).

The Performance Vector in Table 13 shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 13, we obtained for HIGH, MEDIUM and LOW skill, a class precision of 75,00%, 100% and 100% respectively, and a class recall of 100%, 100% and 66,67% respectively, resulting in an accuracy of 88,89%.

**Case 3A:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 4, we obtained the DT presented in Figure 7 and Table 14.

**Figure 7 -** DT referring to Case 3A.



Source: Authors (2021).

According to Figure 7, Rules 1 and 2 define the control variables x4 and x1 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (6) and (7).

$$x4 > 1.600 \tag{6}$$

$$x4 \leq 1.600 \ \rightarrow x1 \leq 1.110 \tag{7}$$

**Table 14 -** Performance Vector 3A.

| | true HIGH | true MEDIUM | true LOW | class precision |
|---|---|---|---|---|
| **pred. HIGH** | 2 | 1 | 2 | 40,00% |
| **pred. MEDIUM** | 0 | 1 | 0 | 100,00% |
| **pred. LOW** | 0 | 0 | 0 | 0,00% |
| **class recall** | 100,00% | 50,00% | 0,00% | |
| **accuracy** | **50,00%** | | | |

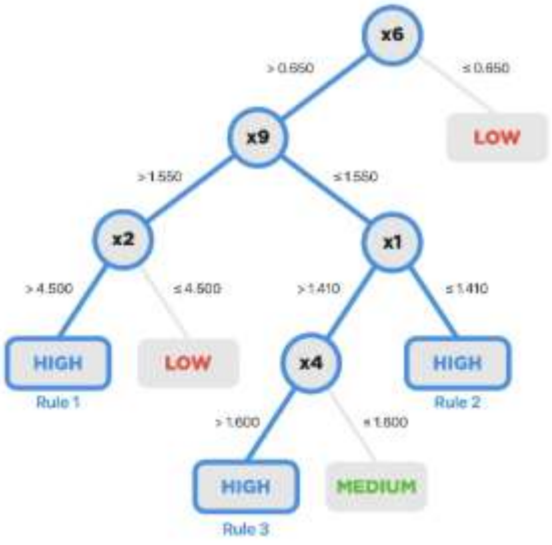
Source: Authors (2021).

The Performance Vector shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 14, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 40%, 100% and 0% respectively, and a class recall of 100%, 50% and 0% respectively, resulting in an accuracy of 50%.

**Case 4A:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 5, we obtained the DT presented in Figure 8 and Table 15.

**Figure 8 -** DT referring to Case 4A.



Source: Authors (2021).

According to Figure 8, Rules 1, 2 and 3 define the control variables x6, x9, x2, x1 and x4 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (8), (9) and (10).

$$x6 > 0.650 \rightarrow x9 > 1.550 \rightarrow x2 > 4.500 \tag{8}$$

$$x6 > 0.650 \rightarrow x9 \leq 1.550 \rightarrow x1 \leq 1.410 \tag{9}$$

$$x6 > 0.650 \rightarrow x9 \leq 1.550 \rightarrow x1 > 1.410 \rightarrow x4 > 1.600 \tag{10}$$

where,

$x_2$ → refers to the distance of Aerodromes;

$x_9$ → refers to Aspect.

**Table 15 -** Performance Vector 4A.

| | true HIGH | true MEDIUM | true LOW | class precision |
|---|---|---|---|---|
| **pred. HIGH** | 1 | 1 | 2 | 25,00% |
| **pred. MEDIUM** | 0 | 1 | 0 | 100,00% |
| **pred. LOW** | 1 | 0 | 0 | 0,00% |
| **class recall** | 50,00% | 50,00% | 0,00% | |
| **accuracy** | **33,33%** | | | |

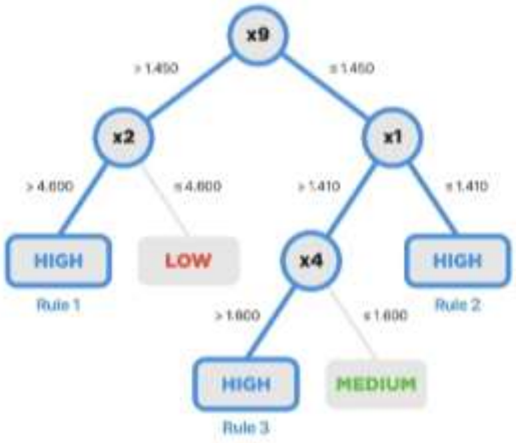
Source: Authors (2021).

The Confusion Matrix in Table 15 shows the Performance Vector, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 15, we obtained for HIGH, MEDIUM and LOW skill, a class precision of 25,00%, 100% and 0,00% respectively, and a class recall of 50%, 50% and 0,00% respectively, resulting in an accuracy of only 33,33%.

**Case 5A:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 6, we obtained the DT presented in Figure 9 and Table 16.

**Figure 9 -** DT referring to Case 3A.



Source: Authors (2021).

According to Figure 9, Rules 1, 2 and 3 define the control variables x9, x2, x1 and x4 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (11), (12) and (13).

$$x9 > 1.450 \rightarrow x2 > 4.600 \qquad (11)$$

$$x9 \leq 1.450 \rightarrow x1 \leq 1.410 \qquad (12)$$

$$x9 \leq 1.450 \rightarrow x1 > 1.410 \rightarrow x4 > 1.600 \qquad (13)$$

**Table 16 -** Performance Vector 5A.

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 2 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 0 | 2 | 0 | 100,00% |
| **pred. LOW** | 0 | 0 | 2 | 100,00% |
| **class recall** | 100,00% | 100,00% | 100,00% | |
| **accuracy** | | **100,00%** | | |

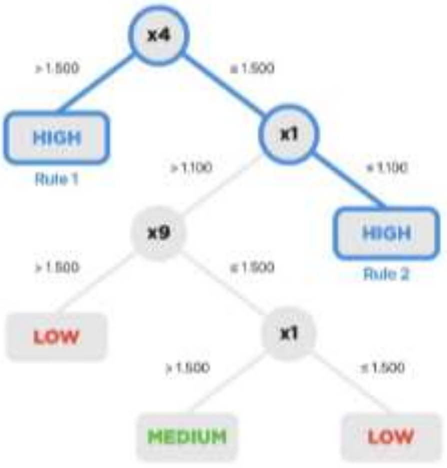
Source: Authors (2021).

The Performance Vector shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 16, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 100% and 100% respectively, and a class recall of 100%, 100% and 100% respectively, resulting in an accuracy of 100%.

**Case 1B:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 7, we obtained the DT presented in Figure 10 and Table 17.

**Figure 10 -** DT referring to Case 1B.



Source: Authors (2021).

According to Figure 10, Rules 1 and 2 define the control variables x4 and x1 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (14) and (15).

$$x4 > 1.500 \tag{14}$$

$$x4 \leq 1.500 \rightarrow x1 \leq 1.100 \tag{15}$$

**Table 17 -** Performance Vector 1B.

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 2 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 0 | 1 | 0 | 100,00% |
| **pred. LOW** | 0 | 1 | 2 | 66,67% |
| **class recall** | 100,00% | 50,00% | 100,00% | |
| **accuracy** | **83,33%** | | | |

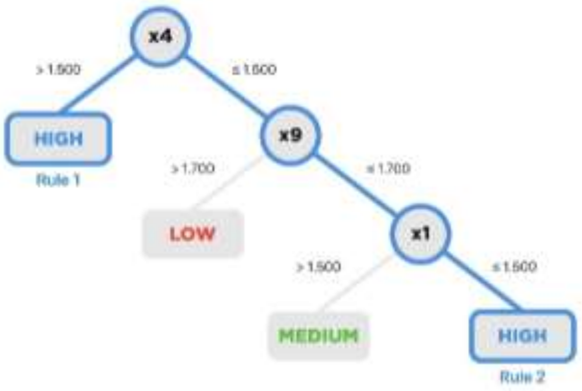
Source: Authors (2021).

The Performance Vector in Table 17 shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 17, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 100% and 66,67% respectively, and a class recall of 100%, 50% and 100% respectively, resulting in an accuracy of 83,33%.

**Case 2B:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 8, we obtained the DT presented in Figure 11 and Table 18.

**Figure 11 -** DT referring to Case 2B.



Source: Authors (2021).

According to Figure 11, Rules 1 and 2 define the control variables x4, x9 and x1 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (16) and (17).

$$x4 > 1.500 \tag{16}$$

$$x4 \leq 1.500 \rightarrow x9 \leq 1.700 \rightarrow x1 \leq 1.500 \tag{17}$$

**Table 18 -** Performance Vector 2B.

| | true HIGH | true MEDIUM | true LOW | class precision |
|---|---|---|---|---|
| **pred. HIGH** | 1 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 0 | 2 | 0 | 100,00% |
| **pred. LOW** | 1 | 0 | 2 | 100,00% |
| **class recall** | 50,00% | 100,00% | 100,00% | |
| **accuracy** | **83,33%** | | | |

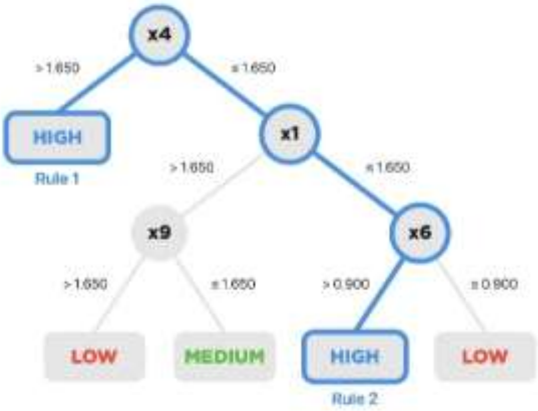
Source: Authors (2021).

The Performance Vector in Table 18 shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 18, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 100% and 100,00% respectively, and a class recall of 50%, 100% and 100% respectively, resulting in an accuracy of 83,33%.

**Case 1C:**

The parameters of the operators used in this case are in their default format, except for the Split Ratio, which used a value of 0.8. After executing the process, with the data presented in Table 9, we obtained the DT presented in Figure 12 and Table 19 and 20.

**Figure 12 -** DT referring to Case 1C.



Source: Authors (2021).

According to Figure 12, which is the same in the results of both validations, Rules 1 and 2 define the control variables x4, x1 and x6 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (13) and (14).

$$x4 > 1.650 \tag{13}$$

$$x4 \leq 1.650 \rightarrow x1 \leq 1.650 \rightarrow x6 > 0.900 \tag{14}$$

**Table 19 -** Performance Vector 1C (Split Validation).

| | true HIGH | true MEDIUM | true LOW | class precision |
|---|---|---|---|---|
| **pred. HIGH** | 3 | 0 | 2 | 60,00% |
| **pred. MEDIUM** | 0 | 5 | 0 | 100,00% |
| **pred. LOW** | 2 | 0 | 2 | 50,00% |
| **class recall** | 60,00% | 100,00% | 50,00% | |
| **accuracy** | | **71,43%** | | |

Source: Authors (2021).

The Performance Vector in Table 19 shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 19, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 60%, 100% and 50,00% respectively, and a class recall of 60%, 100% and 50% respectively, resulting in an accuracy of 71,43%.

**Table 20 -** Performance Vector 1C (Cross Validation).

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 11 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 0 | 12 | 0 | 100,00% |
| **pred. LOW** | 1 | 0 | 9 | 90,00% |
| **class recall** | 91,67% | 100,00% | 100,00% | |
| **accuracy** | **96,97% +/- 10,05%** | | | |
| **Micro Average** | **96,97%** | | | |

Source: Authors (2021).

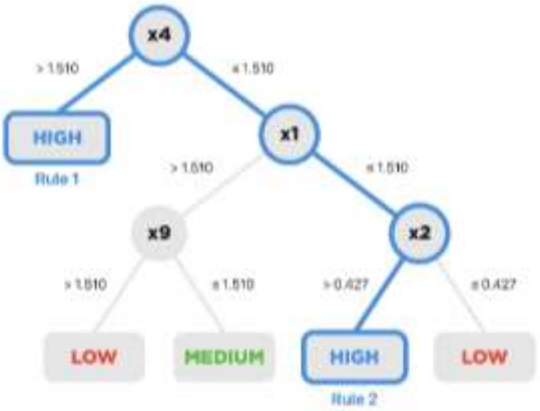
The only modified value among the Cross Validation parameters was the number of folds, changed from 10 to 11. It is possible to observe that Table 20 has more prediction values in all 3 aptitudes, when compared to Table 19. The explanation for this fact is the method of how this validation technique partitions the data into k equal sized subsets and uses only one of them for testing.

The Performance Vector in Table 20 shows a Confusion Matrix with Cross Validation, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 20, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 100% and 90,00% respectively, and a class recall of 91,67%, 100% and 100% respectively, resulting in an accuracy of 96,97% +/- 10,05% and Micro Average of 96,97%. It is interesting to note that the Performance Vector using the Cross Validation has more prediction samples than the Split Validation, this is due to the partitioned into k subsets of equal size used by the Validation method.

**Case 2C:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 10, we obtained the DT presented in Figure 13 and Table 21 and 22.

**Figure 13 -** DT referring to Case 2C.



Source: Authors (2021).

According to Figure 13, which is the same in the results of both validations, Rules 1 and 2 define the control variables x4, x1 and x2 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (20) and (21).

$$x4 > 1.510 \tag{20}$$

$$x4 \leq 1.510 \rightarrow x1 \leq 1.510 \rightarrow x6 > 0.427 \tag{21}$$

**Table 21 -** Performance Vector 2C (Split Validation).

| | true HIGH | true MEDIUM | true LOW | class precision |
|---|---|---|---|---|
| **pred. HIGH** | 3 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 0 | 3 | 0 | 100,00% |
| **pred. LOW** | 3 | 0 | 2 | 100,00% |
| **class recall** | 50,00% | 100,00% | 100,00% | |
| **accuracy** | | **100,00%** | | |

Source: Authors (2021).

The Performance Vector in Table 21 shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 21, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 100% and 100% respectively, and a class recall of 50%, 100% and 100% respectively, resulting in an accuracy of 100%.

**Table 22 -** Performance Vector 2C (Cross Validation).

| | true HIGH | true MEDIUM | true LOW | class precision |
|---|---|---|---|---|
| **pred. HIGH** | 15 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 1 | 16 | 0 | 94,12% |
| **pred. LOW** | 0 | 0 | 12 | 100,00% |
| **class recall** | 93,75% | 100,00% | 100,00% | |
| **accuracy** | | **96,73% +/- 7,54%** | | |
| **Micro Average** | | **97,63%** | | |

Source: Authors (2021).

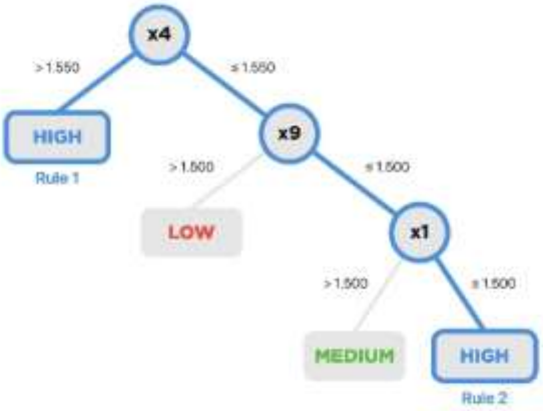
In Case 2C, the only parameter value modified from the Cross Validation pattern was the number of folds, changed from 10 to 11. It is possible to observe that Table 22 has more prediction values in all 3 aptitudes, when compared to Table 21. The explanation for this fact is given by the method of how this validation technique partitions the data between training and testing values.

The Performance Vector in Table 22 shows a Confusion Matrix with Cross Validation, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 22, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 94,12% and 100% respectively, and a class recall of 93,75%, 100% and 100% respectively, resulting in an accuracy of 97,63% +/- 7,54% and Micro Average of 97,63%. It is interesting to note that the Performance Vector using the Cross Validation has more prediction samples than the Split Validation, this is due to the partitioned into k subsets of equal size used by the Validation method.

**Case 3C:**

The parameters of the operators used in this case are in their default format. After executing the process, with the data presented in Table 11, we obtained the DT presented in Figure 14 and Table 23 and 24.

**Figure 14 -** DT referring to Case 3C.



Source: Authors (2021).

According to Figure 14, which is the same in the results of both validations, Rules 1 and 2 define the control variables x4, x9 and x1 in descending order of their degrees of importance. The HIGH prediction skill is highlighted in the figure, and its rules are described, respectively, by expressions (22) and (23).

$$x4 > 1.550 \tag{22}$$

$$x4 \leq 1.550 \rightarrow x9 \leq 1.500 \rightarrow x1 \leq 1.500 \tag{23}$$

**Table 23 -** Performance Vector 3C (Split Validation).

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 2 | 0 | 0 | 100,00% |
| **pred. MEDIUM** | 0 | 2 | 0 | 100,00% |
| **pred. LOW** | 0 | 0 | 2 | 100,00% |
| **class recall** | 100,00% | 100,00% | 100,00% | |
| **accuracy** | | **100,00%** | | |

Source: Authors (2021).

The Performance Vector in Table 23 shows a Confusion Matrix, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 23, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 100%, 100% and 100% respectively, and a class recall of 50%, 100% and 100% respectively, resulting in an accuracy of 100%.

**Table 24 -** Performance Vector 3C (Cross Validation).

| | **true HIGH** | **true MEDIUM** | **true LOW** | **class precision** |
|---|---|---|---|---|
| **pred. HIGH** | 7 | 0 | 1 | 87,50% |
| **pred. MEDIUM** | 1 | 8 | 0 | 88,89% |
| **pred. LOW** | 0 | 0 | 5 | 100,00% |
| **class recall** | 87,50% | 100,00% | 83,33% | |
| **accuracy** | **92,86% +/- 18,16%** | | | |
| **Micro Average** | **90,91%** | | | |

Source: Authors (2021).

To achieve the values in the table above, a single value of the default Cross Validation parameters was modified, the number of folds was changed from 10 to 14. It is possible to observe that Table 24 has more prediction values in all 3 skill, when compared to Table 23. The explanation for this fact is given by the method of how this validation technique partitions the data for the application of its sub-processes.

The Performance Vector in Table 24 shows a Confusion Matrix with Cross Validation, this table helps understand the correct and incorrect prediction made by the algorithm, and shows the accuracy, individual class precision and recall. According to Table 24, we obtained for HIGH, MEDIUM and LOW skills, a class precision of 87,50%, 88,89% and 100% respectively, and a class recall of 87,50%, 100% and 83,33% respectively, resulting in an accuracy of 92,86% +/- 18,16% and Micro Average of 90,91%. It is interesting to note that the Performance Vector using the Cross Validation has more prediction samples than the Split Validation, this is due to the partitioned into k subsets of equal size used by the Validation method.

## 5. Conclusion

The initial impulse for this work came from the current need to assist managers in choosing suitable location to build landfills, as this process involves the analysis of various characteristics of the region. This screening is extremely important to minimize social, environmental and economic impacts in regions that surrounds landfills.

Based on the characteristics of the problem, we chose to work with decision trees, a supervised learning technique for classification, as it is a simple method that builds a predictive structure considering the most relevant attributes for the problem.

All decision trees showed in this article were implemented by using RapidMiner, a data science and artificial intelligence software. As it is a properly validated tool in the business and scientific environment, which presents a collection of algorithms which can be easily implemented and remodeled through block programming, several tests could be performed considering different parameters.

This article also showed the use of arithmetic mean, geometric mean and standard deviation to create extra scenarios for the database, since the initial amount of data proved to be insufficient to work with decision trees.

The results obtained with several tests, considering different numbers of scenarios and parameters in the computational modeling, showed promising results with excellent accuracy and precision. Validating, from a theoretical and statistical point of view, the use of the decision tree to assist in choosing a suitable place for the construction of landfills.

The pursuit of this work is aimed at the application of other Bioinspired techniques. It is expected that, with the implementation of Unsupervised machine learning, such as Clustering and Autoencoders, make it possible to identify other intrinsic characteristics in the source scenarios. Once clustering joins the scenarios according to their similar characteristics, disregarding labeling, and autoencoders try to replicate the input data, perhaps enabling the creation of new scenarios due to the noise present in the process of meeting and decoding of this method.

## Acknowledgments

To the Estácio Belém College for its instruction and structure, and to my undegraduation Advisor Larissa Luz Gomes and to the Federal Institute of Education, Science and Technology of Pará (IFPA) Advisor Denis Carlos Lima Costa from whom i received guidance and support to keep on my academical Journey.

Also, to the research group GM²SC (Gradient of Mathematical Modeling and Computational Simulation) for encouraging research and innovation, and an especial thanks for the coauthors of this wok for the opportunity to be part of the research that result in this paper.

## References

ABNT, N. 13.896 (1997) Aterros de resíduos não perigosos–Critérios para projeto, implantação e operação. https://www.abntcatalogo.com.br/norma.aspx?ID=4829.

Andrade, A. J. B., & Barbosa, N. P. P. (2015). Combinação do método AHP e SIG na seleção de áreas com potenciais para a instalação de aterro sanitário: caso da ilha do Fogo, na República de Cabo Verde. *Revista de Geografia (UFPE)*, *32*(2), 248-266.

Brito, D. A. C., Seabra, L. C., Lima, P. D. M., & Souza, C. M. N. (2020). Manejo De Resíduos Sólidos E De Águas Pluviais: O (Des) Controle Social Em Belém, Pará. *Revista Eletrônica de Gestão e Tecnologias Ambientais*, *8*(2), 103-118.

Chalmers, A. (1999). What is This Thing Called Science? Open University Press.

Crump, T. (2002). A Brief History of Science, as Seen Through the Development of Scientific Instruments. Robinson.

Costa, D. C. L., de Oliveira Costa, H. A., Castro, A. P. S., Cruz, E. C., Neto, J. L. A., & da Cruz, B. C. C. (2020). As dimensões das Modelagens Matemática e Computacional prescrevidas à Gestão Ambiental. *Research, Society and Development*, *9*(10), e6939109013-e6939109013. 10.33448/rsd-v9i10.9013. Retrieved from: https://rsdjournal.org/index.php/rsd/article/view/9013.

Costa, D. C., Nunes, M. V., Vieira, J. P., & Bezerra, U. H. (2016). Decision tree-based security dispatch application in integrated electric power and natural-gas networks. *Electric Power Systems Research*, *141*, 442-449.

de Oliveira Costa, H. A., Costa, D. C. L., & de Meneses, L. A. (2021). Interdisciplinarity Applied to the Optimized Dispatch of Integrated Electricity and Natural Gas Networks using the Genetic Algorithm. *Research, Society and Development*, *10*(2), e42110212641-e42110212641. 10.33448/rsd-v10i2.12641. Retrieved from: https://rsdjournal.org/index.php/rsd/article/view/1264

Crepaldi, P. G., Avila, R. N. P., de Oliveira Paulo, J. P. M., Rodrigues, R., & Martins, R. L. (2011). Um estudo sobre a árvore de decisão e sua importância na habilidade de aprendizado. https://www.inesul.edu.br/revista/arquivos/arq-idvol_15_1320100263.pdf

Freddo, A. R., Nishiyama, M. F., Zanuzo, K., & Koehnlein, E. (2019). Árvores de Decisão como Método de Mineração de Dados: Análise de Prontuários de uma Clínica Escola de Nutrição. *Revista Da Associação Brasileira De Nutrição-RASBRAN*, *10*(2), 31-37.

De Felice, F., Crocetti, D., Parisi, M., Maiuri, V., Moscarelli, E., Caiazzo, R., & Tombolini, V. (2020). Decision tree algorithm in locally advanced rectal cancer: an example of over-interpretation and misuse of a machine learning approach. *Journal of cancer research and clinical oncology*, *146*(3), 761-765.

Garcia, S. C. (2003). O uso de árvores de decisão na descoberta de conhecimento na área da saúde. Rio Grande do Sul: Universidade Federal do Rio Grande do Sul. Retrieved from: http://hdl.handle.net/10183/4703

Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., & Sarker, K. U. (2018, August). Student academic performance prediction by using decision tree algorithm. In *2018 4th international conference on computer and information sciences (ICCOINS)* (pp. 1-5). IEEE.

Yazdi, S., Vosoogh, A., & Bazargan, A. (2018). The Application of Membrane Bioreactors (MBR) for the Removal of Organic Matter, Nutrients, and Heavy Metals from Landfill Leachate.

Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., Ashley, E., & Dudley, J. T. (2018). Artificial Intelligence in Cardiology. *Journal of the American College of Cardiology*, *71*(23), 2668-2679.

Khorram, A., Yousefi, M., Alavi, S. A., & Farsi, J. (2015). Convenient landfill site selection by using fuzzy logic and geographic information systems: a case study in Bardaskan, East of Iran. *Health Scope*, *4*(1) 1-10.

Krestinskaya, O., & James, A. P. (2016, September). Bioinspired memory model for HTM face recognition. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1528-1532). IEEE.

Kumar, A., Sah, B., Singh, A. R., Deng, Y., He, X., Kumar, P., & Bansal, R. C. (2017). A review of multi criteria decision making (MCDM) towards sustainable renewable energy development. *Renewable and Sustainable Energy Reviews*, *69*, 596-609.

Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018). Brain intelligence: go beyond artificial intelligence. *Mobile Networks and Applications*, *23*(2), 368-375.

Magalhães, D. F. V., da Cunha Vieira, M. I. M., & de Souza Norberto, A. (2020). Dimensionamento de geossintético para reforço de aterro sobre solo mole. *Research, Society and Development*, *9*(8), e355985323-e355985323.

Martins, M. E. G. (2013). Desvio padrão amostral. *Revista de ciência elementar*, *1*(1) 022

Mayer, R. E. (2019). Computer games in education. *Annual review of psychology*, *70*, 531-549.

Mu, Y., Liu, X., & Wang, L. (2018). A Pearson's correlation coefficient based decision tree and its parallel implementation. *Information Sciences*, *435*, 40-58.

Moreira, L. L., Schwamback, D., Corrêa, N. R., & Coelho, A. L. N. (2016). SIG Aplicado à seleção de áreas potenciais para instalação de aterro sanitário no município de serra–ES. *Geosciences= Geociências*, *35*(4), 531-541.

Pereira, A. S., Shitsuka, D. M., Parreira, F. J. & Shitsuka, R. (2018). *Metodologia da pesquisa científica. [*UFSM. https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1

Pinheiro, M. M. F., Osco, L. P., Mendes, T. S. G. & Ramos, A. P. M. (2019). Caracterização das áreas restritas para implantação de aterro sanitário na região do Pontal do Paranapanema - SP. In: *Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto*, Santos. São José dos Campos, INPE, 2019. https://proceedings.science/sbsr-2019/papers/caracterizacao-das-areas-restritas-para-implantacao-de-aterro-sanitario-na-regiao-do-pontal-do-paranapanema---sp?lang=en.

Portella, M. O., & Ribeiro, J. C. J. (2014). Aterros sanitários: Aspectos gerais e destino final dos resíduos. *Revista Direito Ambiental e Sociedade, 4*(1), 115-134.

Priya, K.S., Burman, I., Tarafdar, A., & Sinha, A. (2018). Impact of Ammonia Nitrogen on COD Removal Efficiency in Anaerobic Hybrid Membrane Bioreactor Treating Synthetic Leachate. *International Journal of Environmental Research*, 13, 59-65.

Paula, J. A. A. de, Faria, Érica V. de, Lima, A. C. P., Vieira Neto, J. L., & Santos, K. G. dos. (2020). Computational simulation of soybean particles flow in a hopper using computational fluid dynamics (CFD) and discrete elements method (DEM). *Research, Society and Development*, 9(8), e448985463. https://doi.org/10.33448/rsd-v9i8.5463

Ramadhan, I., Sukarno, P., & Nugroho, M. A. (2020, June). Comparative Analysis of K-Nearest Neighbor and Decision Tree in Detecting Distributed Denial of Service. In *2020 8th International Conference on Information and Communication Technology (ICoICT)* (pp. 1-4). IEEE.

RAPIDMINER 9. *Operator Reference Manual*. All rights reserved. RapidMiner GmbH. www.rapidminer.com.

Sathiyanarayanan, P., Pavithra, S., Saranya, M. S., & Makeswari, M. (2019, March). Identification of breast cancer using the decision tree algorithm. In 2019 *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1-6). IEEE.

Şener, Ş., Şener, E., Nas, B., & Karagüzel, R. (2010). Combining AHP with GIS for landfill site selection: a case study in the Lake Beyşehir catchment area (Konya, Turkey). *Waste management*, *30*(11), 2037-2046.

Souto, G. D. (2009). Lixiviado de aterros sanitários brasileiros: estudo de remoção do nitrogênio amoniacal por processo de arraste com ar (stripping) (Doctoral dissertation, Universidade de São Paulo).

Sodré, G. R. C., de Freitas, S. J. N., Rodrigues, J. B., Igawa, T. K., de Sousa Amorim, I. L., & Cabral, A. C. L. C. (2020). Avaliação sustentável para instalação de aterro sanitário em uma cidade da Amazônia oriental. *Nature and Conservation*, 13(3), 112-121.

Swacha, J., Maskeliūnas, R., Damaševičius, R., Kulikajevas, A., Blažauskas, T., Muszyńska, K., & Kowalska, M. (2021). Introducing Sustainable Development Topics into Computer Science Education: Design and Evaluation of the Eco JSity Game. *Sustainability*, 13(8), 4244.

Szczepanski, M. (2019). Economic impacts of artificial intelligence (AI). *European Parliamentary Research Service* (PE 637.967).

Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), 337-339.

Yang, X., Zhou, T., Zwang, T. J., Hong, G., Zhao, Y., Viveros, R. D., & Lieber, C. M. (2019). Bioinspired neuron-like electronics. *Nature materials*, 18(5), 510-517.

Yoo, S. H., Geng, H., Chiu, T. L., Yu, S. K., Cho, D. C., Heo, J., & Lee, H. (2020). Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Frontiers in medicine*, 7, 427.

Wu, Y., Zhou, J., Hu, Y., Li, L., & Sun, X. (2018). A TODIM-based investment decision framework for commercial distributed PV projects under the energy performance contracting (EPC) business model: A case in East-Central China. *Energies*, 11(5), 1210.