Estimated Maintenance Costs of Brazilian Highways Using Machine Learning Algorithms

Ricardo Gaussmann [1]*, Dennis Coelho [1], Anita Fernandes [1,2], Paul Crocker [3,4], Valderi R. Q. Leithardt [2,4,5]

[1] University of the Itajai Valley, Specialization Course in Big Data, Santa Catarina, BRAZIL
[2] University of the Itajaí Valley, Master in Applied Computing, Santa Catarina, BRAZIL
[3] Telecommunications Institute, IT Branch, Covilha, PORTUGAL
[4] Department of Informatics, University of Beira Interior, Covilha, PORTUGAL
[5] COPELABS, Lusophone University of Humanities and Technologies, Lisboa, PORTUGAL
***Corresponding Author:** ricardog@softplan.com.br

ARTICLE INFO	ABSTRACT
Published: 30 Jul. 2020	The road infrastructure is considered to be a key prerequisite of social and economic development of any country and therefore solutions that assist in the management and maintenance of this key infrastructure are important. This paper presents the application of Machine Learning algorithms, such as Multilayer Perceptron Neural Network and K-means for estimating the level of services required for highway conservation in Brazil. The data used is from the Federal District highways, recorded in the form of Service Orders in the Road Administration System, as well as the road solutions catalog elaborated from the price table of the Federal District Roads Department. A database was created containing data for routine maintenance history, road solutions catalog and price lists. The machine learning algorithms were applied and evaluated, and it was concluded that the K-means algorithm had the best performance for estimating the maintenance costs of Brazilian highways. **Keywords:** k-means, multilayer perceptron, machine learning, road maintenance costs

INTRODUCTION

According to Ivanova and Masarova (2013), since the importance of the road network transcends national boundaries, the expansion and upgrade of the road network is vital to increase economic performance. Hence, poor road infrastructure poses a hindrance to investments in countries depending on them in terms of their economic performance and competitiveness enhancement. Infrastructure investments impact a country's economy through direct channels, such as expanding supply capacity or production outlets; and indirect channels, such as the improvement of total factor productivity, providing for the nation's economic and social development. With regard to transport infrastructure, in Brazil, the main focus of investments is the highways. **Figure 1** presents Brazil's traffic flow of highways from Jan 1999 to Jun 2019, according to CEICDATA (2020), and **Figure 2** presents Brazil's traffic flow of highways considering only heavy vehicles from Jan 1999 to Jun 2019, according to CEICDATA (2020).

According to Oladele, Adedimila and Egwurube (2011), it has been estimated that the trend in annual expenditure on road maintenance in developing countries had increased. Therefore, as resources are dwindling, the challenge of highway engineering infrastructure is an economic one, demanding increasing emphases on the optimum allocation and use of these scarce resources. The road infrastructure comprises all types of roads in a given area, including various other structures, that serves to transport passengers and goods. The road infrastructure includes all road categories, facilities, structures, signage and markings, electrical systems, and so on needed to provide for safe, trouble-free and efficient traffic (Ivanova and Masarova, 2013).

Transport infrastructure is essential in determining a country's per capita income level, as there is a direct relationship between paved road availability and human development (OECD, 2017). The Brazilian Institute for Applied Economic Research has shown results that public investment in the transportation sector has a positive and statistically significant effect on long-term economic performance in Brazilian states, and potentially contributes to the reduction of income inequality among them (Wolff and Caldas, 2018).

According to Amann et al (2016), millions of reais are invested annually in the maintenance and conservation of the Brazilian federal highways, which make up the National Road Plan, and of the state highways belonging to the States Road Systems.
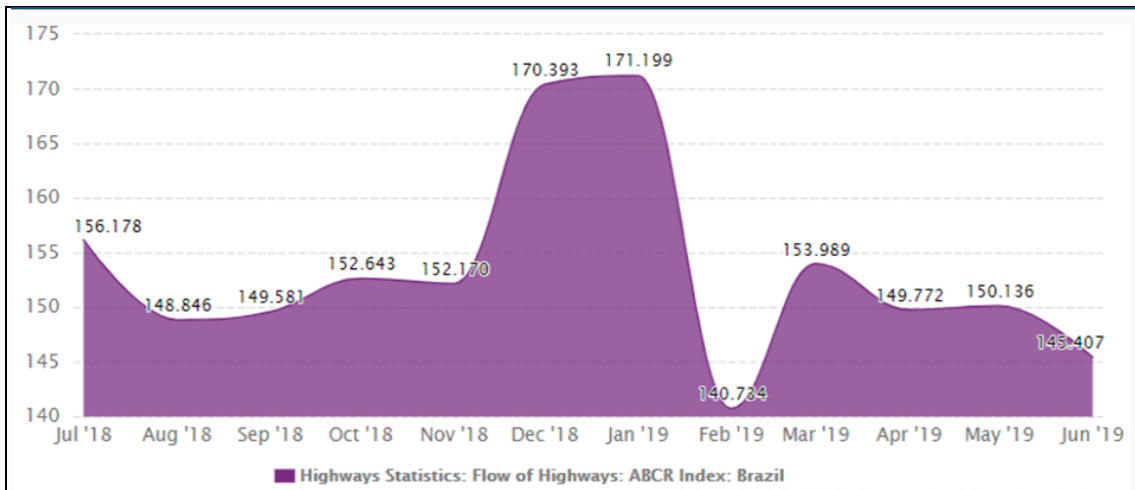
Figure 1. Brazil's flow of highways from Jan 1999 to Jun 2019 (CEICDATA, 2020)
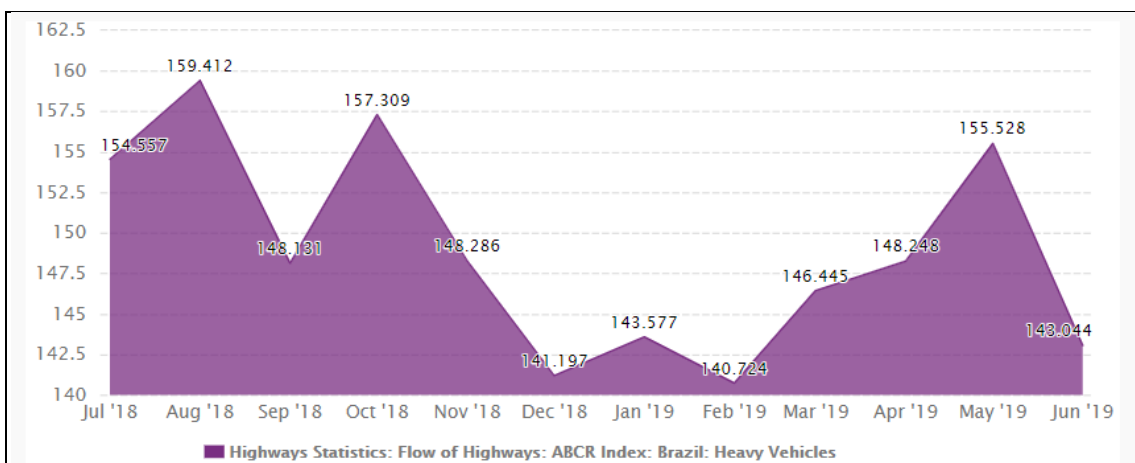

Figure 2. Brazil's flow of highways considering only heavy vehicles from Jan 1999 to Jun 2019 (CEICDATA, 2020)

Mourougane and Pisu (2011) presented in their studies, showed that it is possible to achieve a 7.8% reduction in average fuel consumption and 18.7% reduction in truck maintenance expenses for vehicles that travel in areas with better conservation conditions. The operating costs of a truck can double in relation to the use of a highway in bad condition to another one in good condition (Amann et al., 2016).

These researches presented results that serve as justification for the high direct and indirect costs that the deficiency of the national road network entails to the Brazilian economic system, and reinforce the need to direct a significant portion of resources for a broad program of road maintenance.

Research with companies that have revenues considered significant in relation to the Brazilian Gross Domestic Product, have shown that 61% consider the recovery and expansion of highways as the most important work to be carried out by the public sector (Wolff and Caldas, 2018). Another relevant information is that 54% of companies that participated in a survey selected poor road conditions as among the biggest factors that could increase their logistics costs (Mourougane and Pisu, 2011).

According to the Brazilian National Transport Confederation (Mourougane and Pisu, 2011), the National Road System of Brazil has a road network of around 1.7 million kilometers, including federal, state and municipal roads, representing the fourth largest road network in the world. However, only about 14% of existing roads are paved. This percentage is close to some Latin American countries, but considering the world scenario, it is very far from. For example, the United States, Japan, China, France and Germany (Ghisolfi et al., 2019).

The Brazilian National Road Plan encompasses the federal highways under the circumscription of the National Department of Transport Infrastructure, which are built and maintained by the National Superintendencies and Local Units of the national agency, located in the state capitals and main municipalities, respectively. The States Road Systems, on the other hand, are the highways under the circumscription of the states and the Federal District. Also part of the States Road Systems are federal highways delegated to the states, through delegation agreements (Amann et al., 2016; Wolff and Caldas, 2018).

Although the privatization and concession of highways is a trend in the Brazilian scenario, there is still a huge network under the jurisdiction of the public sector. Road concessions have great autonomy and resources available to make the necessary investments. These scenarios, quite opposite to those of public authorities which face severe budgetary constraints, must comply with the legal criteria for hiring specialized companies to perform the services of road maintenance and conservation (Gaussmann et al., 2020; Pivatto et al., 2017).

In order to regulate the legal criteria for public procurement in Brazil, Law 8,666 was sanctioned in 1993, which establishes that the procurement and acquisition of goods and services in the public sector must be carried out through the elaboration of a detailed budget of the desired object, and based on the reference price table adopted by the agency, the overall value of the acquisition or contracting is obtained (Pivatto et al., 2017).

Within this context, the Department of Highways of the Brazilian Federal District implemented in 2016 a computer system called Road Maintenance and Conservation Management System to control and plan the maintenance and conservation activities of highways (Gaussmann et al., 2020).

The Road Maintenance and Conservation Management System aims to maintain an updated register of road elements, to qualify the state of conservation of each road segment and its elements, to provide subsidies for the quantification of services and resource needs for conservation and budget proposals; the system provides planning and management of services and ensures the control of services performed (Martin, 2018). To estimate road maintenance costs the Road Maintenance and Conservation Management System uses "effort-based calculation" to answer how much conservation should be performed (Gaussmann et al., 2020).

It is intuitive that the higher the level of effort, the greater the amount of conservation and the higher the cost. Thus, for example, for manual weeding of an unpaved shoulder, an effort level of 0.5, 1.0, or 2.0 could be defined for each square meter of unpaved shoulder, which is equivalent to weeding once every two years, once a year or twice a year respectively by applying the three exemplified rates. It is also evident, on the other hand, that the higher the level of effort applied, the higher the level of quality achieved (Gaussmann et al., 2020).

The example presented is intended to explain what is meant by level of effort, as a formal definition would not be very clear. We can say what level of effort is the amount of conservation service (in the case of the example weeding) to be performed per unit of generator element (m^2 of unpaved shoulder) per year to produce or reach a certain level of quality. Thus, the assumption of a certain level of quality, a subjective quantity is implicit in the value attributed to the level of effort.

The subjective aspect of this analysis generates many biases and distorted analyzes. Consequently, it leads to damage to the public coffers.

In this sense, some studies have presented the use of Machine Learning algorithms to assist in the design, cost analysis and maintenance of the roads, such as the studies presented in (Alfakis, Alatawi and Abushandi, 2014; Cao, 2019; Huseynov, Ribera and Palma, 2018; Piryonisi and El-Diraby, 2017). In the next section, we review other related work that illustrate the use of machine learning algorithms.

Machine Learning Applications

Wilmot and Mei (2005) presented the development of a procedure that estimates the escalation of highway construction costs over time, using an artificial neural network model which relates overall highway construction costs, described in terms of a highway construction cost index, to the cost of construction material, labor, and equipment, the characteristics of the contract and the contracting environment prevailing at the time the contract was let. Results demonstrate that the model is able to replicate past highway construction trends in Louisiana with reasonable accuracy.

Sodikov (2005) presented a research that used an artificial neural network approach. The data were from ROCKS (Road Costs Knowledge System). This database contains road works cost data from 65 developing countries. Poland and Thailand have a relatively large number of projects. For this reason, they were chosen to investigate the relationship between project cost and other variables such as work activity, terrain type, road parameters, etc. The artificial neural network approach used was Multilayer Perceptron. Comparing artificial neural network with multiple regression model, the neural network used had a better accuracy.

Chou (2009) showed that with the growth of transportation networks in developing countries, the cost-efficacy control of maintenance operations has become critical to the infrastructure asset management after highway construction. In this context, the efficient management of numerous annual projects with limited resources needs to accurately estimate costs and leave a trail of project information during the process of making maintenance project selection decision. So, a Case- Based Reasoning expert prototype was developed to compare historical data at the work item-level across the case library in order to determine preliminary project cost with readily available information rapidly based on previous experience of pavement maintenance related construction to assist decision makers in project screening and budget allocation. The analytical results demonstrate the ability of the system to estimate the item-level cost of pavement maintenance projects with the satisfactory precision during the conceptual project phase.

Cao, Ashuri and Baek (2018) explained that resurfacing is one the most common highway projects in Georgia and constitutes a large portion of the state´s highway investment every year. The value of the unit price bid is one of the leading indicators to comprehensively reflect the cost to the Georgia Department of Transportation for these projects. The authors proposed a robust ensemble learning model to predict the value of unit price bids. The results are compared with those from a baseline Monte Carlo simulation and multiple linear regression model. Comparison shows that the proposed ensemble learning model performs much better than any single machine learning model and the baseline models.

Barros, Marcy and Carvalho (2018), presented the development of an estimation technique for construction highway projects using artificial neural networks. Different architectures of the network with 10, 15 and 20 neurons were trained and tested with the backpropagation algorithm. Based on this, data from fourteen highway projects in Brazil were collected and analyzed. Eleven parameters that contribute the most of the construction final budget were found after trials and errors. For the best scenario, an

average cost estimation accuracy of 99% was achieved. This preliminary study showed the feasibility of the artificial neural network applied to projects in Brazil and may be used by public agencies in the future.

In the next section we present concepts related to Machine Learning, Artificial Neural Networks, and K-means.

Machine Learning, Artificial Neural Networks, and K-means Concepts

Machine Learning (ML) is a form of Artificial Intelligence that enables a system to learn from data rather than through explicit programming (Hurwitz and Kirsch, 2018). ML uses a variety of algorithms that iteratively learn from data to improve, describe data, and predict outcomes.

ML has a large overlap with statistics, as both areas study data analysis. However, unlike statistics, which focuses mainly on well-defined theoretical models and parameter adjustments to these models, ML has a more algorithmic focus, using more flexible and heuristic model representations to perform the search. For example, a statistical analysis may determine distributions, covariances, and correlations among the attributes that describe the facts, but it is not able to characterize these dependencies on an abstract and conceptual level as humans do; nor provide a causal explanation of why these dependencies exist (Hurwitz and Kirsch, 2018).

While a statistical analysis of the data can determine the central trends and variances of certain factors, it cannot produce a qualitative description of the regularities, nor can it determine the dependencies on factors not explicitly provided with the data (Huseynov, Ribera and Palma, 2018).

The flexibility of ML methods enables learning through data that has not been collected from a rigorous controlled experimental process but obtained from any process whose primary purpose is not the discovery of knowledge (Shalev-Shwartz and Ben-David, 2014).

Considering the literature previously presented, artificial neural networks and k-means have been widely used in research on costs in highway design. Monte Carlo simulation was also used in some studies. In this context, for the research presented here, they will be used.

Artificial Neural Networks refer to a biologically inspired computational model consisting of simple processing elements (artificial neurons) that apply a certain mathematical function to the data (activation function) generating a single response, are layered and linked together, these connections being generally associated with coefficients called weights. The adjustment of these weights is performed by a process called training or learning, being responsible for the extraction of data characteristics and knowledge storage of networks (Mitchell, Carbonell and Michalski, 2016).

K-means uses the data grouping algorithm by means. The goal is to find the best division of P given in k groups C_i, i = 1, …, k, so that the total distance between the data of a group and its center, summed by all groups, is minimized. According to Kelleher, MacNamee and D´Arcy (Bishop, 2016), k-means uses the values of the first n cases in a data file, with temporary estimates of k-clusters averages, where k is the user-specified number of clusters. Thus, the center of the initial cluster id formed for each case around the nearest data and then compared to the most distant points and the other clusters formed. From there, within a process of continuous updating and interactive process are the centers of the final's clusters (Kelleher, Namee and D´Arcy, 2015).

In other words, the algorithm randomly assigns P points to k groups and averages the vectors of each group. Then each point is shifted to the group corresponding to the average vector to which it is closest. With this new rearrangement the points in k groups, new mean vectors are calculated. The process of relocating points to new groups whose average vectors are closest to them continues until a situation is reached where all points are already in the groups of their nearest average vectors (Bonaccorso, 2017).

Based on these two Machine Learning algorithms, we sought to estimate the road maintenance costs of the Federal District.

The following section will be presented as calculating the estimate from the data collected and stored in the Road Maintenance and Conservation Management System database, and how was the calculation calculated using the Machine Learning algorithms.

ESTIMATION OF SERVICES QUANTITIES USING ROAD MAINTENANCE AND CONSERVATION MANAGEMENT SYSTEM

In accordance with Article 2nd from Brazilian Law 8.666 of June 21, 1993, "Public Administration works, services, including advertising, purchases, disposals, concessions, permits and leases, when contracted with third parties, will necessarily be preceded by bidding."

In order for public agencies, including those related to the highways, to have a reference price for the works, maintenance and conservation services of their highways, reference price tables are prepared by the agency itself, as well as auxiliary tables, such as the Referential Costs of Works System table prepared by the National Department of Transport Infrastructure, and the National Costs and Indexes Research System table, prepared by the state bank called Caixa Econômica Federal (Raschka, 2015).

The System Costs and Budgets of Works, used by the Federal District Department of Roads (DER/DF), allows the registration of various price lists in force, they are differentiated by the nature of the work, such as construction of highways, among others. This system communicates with the Road Maintenance and Conservation Management System, which establishes which indicative price and service table will be used as the basis for its calculations. For this research we used the Road Maintenance and Conservation Management System Referential Table. **Table 1** presents examples of some services used in the study.

Table 1. DER/DF Pricing Table Services - May 2018

ID	Description	Unit	Unit Cost R$
3 S 08 100 00	Hole Cover with hot machined bituminous concrete	m³	282,37
2 S 02 540 01	Supply hot machined bituminous concrete	T	77,98
3 S 08 400 00	Signpost Cleaning	m²	7,09
3 S 08 302 01	Manhole Cleaning	m³	18,19
3 S 08 300 01	Gutter Cleaning	m	0,54
3 S 08 302 01 #1	Gallery Cleaning	m³	18,19
3 S 08 901 01	Lawn mowing and cleaning	m²	0,13
4 S 06 200 02 #1	Signposting	m²	439,30
4 S 06 200 91 #1	Signpost removal	m²	26,44

Table 2. Pavement Type

Class	Description
1	Porland Cement Concrete
2	Superficial Treatment
3	Primary Coating
4	Bituminous Concrete
5	Other Surfaces
6	Coating Independent

Table 3. Details of Effort Levels for the Hole Covering Service

Level	Element Type	Pavement Class	Traffic	Effort Level	Unit
1	Rolling Track and Shoulder	Bituminous Concrete	3001 to 5000	.840	m³/km
2	Rolling Track and Shoulder	Bituminous Concrete	3001 to 5000	.520	m³/km
3	Rolling Track and Shoulder	Bituminous Concrete	3001 to 5000	.240	m³/km
4	Rolling Track and Shoulder	Bituminous Concrete	3001 to 5000	.120	m³/km
5	Rolling Track and Shoulder	Bituminous Concrete	3001 to 5000	.00	m³/km

The current methodology used by the DER/DF for the elaboration of quantitative maintenance and conservation services of the road network for estimating cost budgets, is called "Effort Levels". The effort levels determine the amount of service according to the condition of the pavements and other road elements (vegetation, signaling, safety and drainage).

For each type of road element, a grade relating to the conservation state is given, with grades between 1 and 5 (grade 1 refers to the worst case and grade 5 refers to the best condition of the feature). The levels of effort, condition, can be mapped as follows: 1- terrible; 2 - bad; 3 - regular; 4 - good; 5 - great.

The types of road elements defined as point or transverse, such as manholes and signposts, are assessed individually, for example, for such elements on the road a single grade is given. For the longitudinal elements, such as metal fenders, drainage ditches, and even the running tracks and hard shoulders, they are divided into segments with varying lengths, according to their conservation grade. It is noteworthy that such field surveys are performed annually by the engineers and technicians of the DER/DF highway districts.

The definition of the effort levels is categorized for each type of road element, in each grade/situation, taking also into consideration the traffic classes of the highways. In the case of pavement services, the classification of the pavement type is also considered, as shown in **Table 2**.

The current methodology used by the DER/DF for the elaboration of quantitative maintenance and conservation services of the road network, consequently estimating these costs budgets, is called "Effort Levels" (**Table 3**). The effort levels determine the amount of service according to the condition of the pavements and other road elements (vegetation, signaling, safety and drainage).

The application of the effort level methodology occurs in a routine of the Road Maintenance and Conservation Management System, performed in each stretch of the road network under the jurisdiction of the agency, in this case the DER/DF.

This routine evaluates all existing elements on each stretch of highway, considering the grade/status of the element, the applicable services, the traffic class and the pavement class, if necessary. Through this information, the level of effort representing the element is identified, and the amount of service is calculated.

The amount of service is the product of multiplying the element length, extension or area by the defined effort level. Quantities of services are performed annually at the beginning of the second half of the year, due to the need to include the financial resources required to perform the services in the next year's budget plan. After calculating the service quantities, the unit costs of the services defined in the DER/DF price reference table are considered, and finally the annual budget is calculated.

The execution of services is controlled in annual cycles, with periods beginning in January and ending in December. Monthly service schedules performed from the issuance of work orders. Each work order can contain several services to be performed on different road sections.

Table 4. Inconsistencies identified in data provided by DER/DF

Inconsistency ID	Detailing
1	Missing or incomplete track note survey information on some road sections.
2	Some road elements have no conservation grade/status.
3	Drainage service execution records in sections that do not have drainage devices in the cadastral inventory.
4	Significant track service execution records on sections that have a regular, good or great track rating.
5	Service release records with wrong units of measure.

ESTIMATION OF SERVICES QUANTITIES USING MACHINE LEARNING

In accordance with Article 2nd from Brazilian Law 8.666 of June 21, 1993, "Public Administration works, services, including advertising, purchases, disposals, concessions, permits and leases, when contracted with third parties, will necessarily be preceded by bidding."

To estimate the amount of services and annual maintenance and conservation budgets of the highways under the DER/DF jurisdiction, in a more effective way, reducing the subjectivity issues present in the approach described above, it was proposed the use of Machine Learning in order to establish the behavioral patterns of the requested services. As a feasibility study of the application of such an approach, it was established that the situation of the running lanes and the other road elements would be based.

In this context, three tabular data sets were extracted from Road Maintenance and Conservation Management System. The first data set refers to the visual assessment carried out on the lane, third-lane and road shoulder pavements, including the situation of the pavements. The second data set contains the inventory of road elements deployed on the highways, whose function is to provide additional features to the highways, such as signaling and safety of their users, as well as to provide rainwater drainage. The third data set has the records of all maintenance and conservation services performed on the road network under the circumscription of the DER/DF.

The Application of Machine Learning

This work used two Machine Learning approaches. The first approach was based on Artificial Neural Networks with Backpropagation learning algorithm. Subsequently, the k-means clustering algorithm was used.

To create the Artificial Neural Network data set, data relating to the evaluation of pavements, road elements and services performed were grouped by road sections in spreadsheets. This way, each section obtained the average track grade and the information of the road elements in a single line of the spreadsheet. The costs of services performed in the road segments (stretches) were also considered, grouped into track services, vegetation services, safety and signaling services (linear elements), safety and signaling services (punctual) and drainage services.

It is noteworthy that the prepared spreadsheet considered all 462 sections under the DER/DF jurisdiction. Relevant additional information was added to this spreadsheet, such as traffic on each stretch and year of construction/restoration of the highways.

For the input data set of the Artificial Neural Network, the following variables were considered: length of stretch; traffic factor (high, medium or low); average track grade; number of bearing tracks; number of shoulders; vegetation area; extension of safety/signaling devices (linear); number of safety/signaling devices (punctual); and extension of drainage devices.

For the output data set, the following variables were considered: cost of services applied to the runway; cost of services applied to vegetation; cost of services applied to security elements (linear); cost of services applied to security elements (punctual); and cost of services applied to drainage devices.

These variables were chosen due to their quality. That is, these variables, when preprocessed, did not suffer significant losses.

The configuration of the Artificial Neural Network used had nine neurons in the input layer and five neurons in the output layer. The intermediate layer was configured with forty neurons.

For the training of the Artificial Neural Network, following what the literature of the area recommends, 70% of the records of the created spreadsheet were selected, and the other 30% were used for testing. It was found in the training and testing of the network that the learning rate was very low. Using this setup, it was necessary to do some tests without input variables, in order to understand the behavior of the network. Variables such as traffic factor, vegetation area and extent of drainage devices were eliminated. The eliminated variables were selected based on consultation with experts in the field, who are DER / DF engineers. None of the trials showed an improvement in the learning rate of the network.

This situation required a more detailed analysis of the data provided by DER/DF, to explain the reason for the low learning rate of the Artificial Neural Network. We analyzed the three data sets regarding the track situation, road element inventory, and maintenance and conservation services that were performed. Five categories of inconsistency were identified, which are listed in **Table 4**.

In addition to identifying problems in the registration of the road network inventory (inconsistency 1), it was also noted the existence of records of erroneous service execution, such as the mowing service of vegetation areas, which is measured in hectares. and the recorded values refer to square meters.

These factors explained the low learning rate of the Artificial Neural Network and made the application of this algorithm unsuitable for the database under analysis. Thus, it became necessary to evaluate alternatives.

As an alternative to the Artificial Neural Network, the k-means algorithm was chosen. For the application of this algorithm, the number of ten clusters was stipulated. The data evaluated were the same as those used by the artificial neural network. After data clustering, analysis of the defined groups was performed. These analyzes consisted of mathematical calculations of mean and standard deviation of each cluster for each of the service groups: runway, vegetation, safety/signaling (linear devices), safety/signaling (point devices) and drainage.

From these calculations, comparative analyzes were performed of values of each of these service groups of each stretch and the respective average of their clusters, obtaining the number of standard deviations of the records.

To compare the patterns found in the clustering of available data sets, we opted for Monte Carlo Simulation. Monte Carlo Simulation is a method that makes use of random, or pseudorandom, numbers to calculate not necessarily random quantities. According to Chen and Chen (2017), Monte Carlo Simulation consists of generating random values for each probability distribution within a model, with the objective of producing hundreds or thousands of scenarios. The distribution of calculated values (for each case) should reflect their probability of occurrence.

In the case of the work in question, one hundred thousand repetitions of calculations were performed, in which samples were randomly selected from fifty records of different sections of the highway in each repetition. Each repetition performed calculations of differences between the individual costs of each road segment, for the services of the lane, vegetation, signaling/safety (linear and point devices) and drainage groups, and the average costs of each cluster.

With this it was possible to calculate the percentages of errors between the costs of the sections and their clusters, for each of the service groups.

Comparing the results of Monte Carlo Simulation, we can see that the error rate was greatly reduced, demonstrating the effectiveness of the method. Considering rolling track, the error decreased 30%. For vegetation, the error decreased 43%, for drainage, 58%. Security (linear) and Security (on point) had had the largest decrease in the error rate, 60%, and 65% respectively.

Based on the Monte Carlo simulation, it was realized that the error regarding each variable can be considered acceptable, this gives a clear evidence that the application of k-means may yield a more satisfactory result.

CONCLUSIONS

Through the studies and research carried out, it is possible to highlight the importance of analyzing investment management methodologies and procedures, especially when they are performed by the government and that provide direct or indirect benefits to society, as these are significant sums of money and needs to be invested wisely in order to maximize results. Another point observed was the lack of consolidated information on data collected in the road network under the agency's circumscription, as some segments do not have such data in whole or in part.

The direct application of an artificial neural networks algorithm for the definition of data patterns for the proposed alternative to the curernt system of analysis used by Road Maintenance and Conservation Management System was found to be inappropriate.

However, the use of a clustering algorithm for grouping similar information proved to be quite applicable to the case study, since the groupings of excerpts made it possible to obtain very assertive service execution patterns. This was confirmed by the histograms generated from the Monte Carlo Simulation, which showed the homogeneity of the analyzed and grouped data (**Figure 3**).

Within this context, it was proposed to the DER/DF the implementation of this new methodology to complement the one currently used, in order to cluster the road sections, to anticipate maintenance and conservation services needs in new road sections that may be incorporated into the road network of the Federal District.

In parallel to this, it was proposed to the technical team of the agency that they evaluate the use of artificial neural networks for service estimates, considering that the inventory of road elements will be kept up to date and complete, and that the database of implementation of the services contain all the work done, and not only for short periods of time.

In order for this to happen various measures have subsequently been taken by the management of the DER/DF. The first was the implementation of data validation routines expected in the registrations, entries and closings of the service orders issued, for the execution of the services on the highways. With this measure, the quantities of materials and tools used, working hours of machines and road implements, as well as the productive and unproductive hours of the employees of the teams involved started to be verified so that quantities above or below those expected are rejected by the system. This has started to provide greater reliability in the road inventory records to the agency.

Another extremely important measure adopted was the development of a new module - called RoadMaps - integrated with the management software used. This new module was designed with two important features. The first, allows teams in the field to receive the data of the orders assigned to them, as well as record services performed and their numbers directly in a mobile application. Regarding the second functionality, it is also a mobile application for conducting road inventories, in which new elements can be registered, and the current elements that already appear in the inventory can be adjusted and their grade / situation reclassified. This module also has web functionality for monitoring and controlling updated inventory.

And last, but not least, was the hiring of a company specialized in carrying out road registration and field survey services but using the new RoadMaps module developed for this purpose.
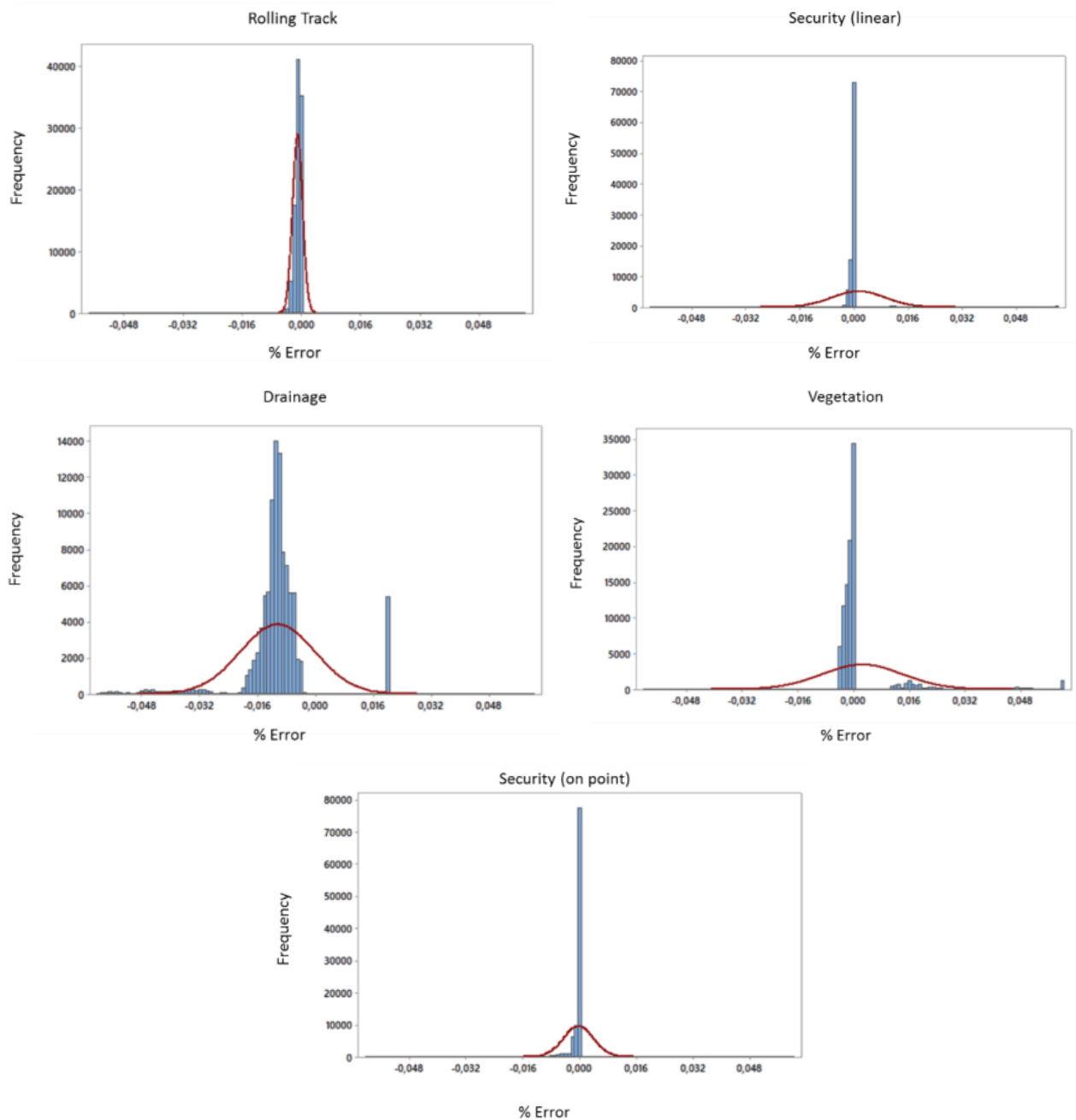
Figure 3. Error rate histogram by Monte Carlo Simulation

From the adoption of these measures, the DER/DF expects to obtain a mass of data that is very consistent both in terms of updated road inventory, as well as for accurate financial accounting for the control of public spending on the highway infrastructure of the Federal District.

To date, the work of field surveys is in progress, and is expected to be completed in August 2020. After completion, this mass of updated data will be used for the application of back propagation algorithms to analyze the initial proposition in this case study, for the determination of behaviors and annual financial predictability.

Until then, the grouping of road stretches using the "K-means" classification algorithms is being adopted as a way of comparing the results obtained for maintenance costing the stretches of highways under the responsibility of the DER/DF.

REFERENCES

Amann, E., Baer, W., Trebat, T. and Villa Lora, J. (2016). Infrastructure and its role in Brazil´s development process. *The Quarterly Review of Economics and Finance*, *62*(November, 2016), 66-73. https://doi.org/10.1016/j.qref.2016.07.007

Barros, L. B., Marcy, M. and Carvalho, M. T. M. (2018). Construction cost estimation of Brazilian highways using artificial neural networks. *International Journal of Structural and Civil Engineering Research*, 7(3), 283-289. https://doi.org/10.18178/ijscer.7.3.283-289

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning*. Springer. ISBN-13: 978-0387310732.

Bonaccorso, G. (2017). *Machine Learning Algorithms*. Packt Publishing, Birmighan. ISBN 1785884514, 9781785884511.

Cao, Y. (2019). *Modeling and predicting the variation of US highway construction cost* (Doctoral Dissertation), School of Building Construction, Georgia Institute of Technlogy. Available at: http://hdl.handle.net/1853/61237 (Accessed: 3 November 2019).

Cao, Y., Ashuri, B. and Baek, M. (2018). Prediction of unit price bids of resurfacing highway projects through ensemble machine learning. *Journal of Computing in Civil Engineering*, 32(5). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000788

CEICDATA (2020). *Brazil Highways Statistics: Flow of Highways: Flow of Vehicles Index*. Available at: https://www.ceicdata.com/en/brazil/highways-statistics-flow-of-highways-flow-of-vehicles-index. (Accessed: 10 July 2020).

Chen, D. and Chen, J. D. (2017) *Monte Carlo Simulation Based Statistical Modeling*. Springer. https://doi.org/10.1007/978-981-10-3307-0

Chou, J. S. (2009) Web based CBR System applied to early cost budgeting for pavement maintenance project. *Expert System with applications*, 36(2), 2947-2960. https://doi.org/10.1016/j.eswa.2008.01.025

Elfaki, A. O., Alatawi, S. and Abushandi, E. (2014). Using intelligent techniques in construction project cost estimation: 10 year survey. *Advances in Civil Engineering*, 2014. Article ID 107926. https://doi.org/10.1155/2014/107926

Gaussmann. R., Coelho, D.K, Fernandes, A., Crocker, P. and Leithardt, V. (2020). Using Machine Learning For Road Maintenance Cost Estimates In Brazil: A Case Study In The Federal District. In 2020 *15th Iberian Conference on Information Systems and Technologies* (CISTI), IEEE Xplore, ISBN: 978-989-54659-0-3. https://doi.org/10.23919/CISTI49556.2020.9141148

Ghisolfi, V., Ribeiro, G. M., Chaves, G. L. D. and Orrico Filho, R. D. (2019) Evaluating impacts of overweight in road freight transportation: a case study in Brazil with system dynamics. *Sustainability*, 11, 3128. https://doi.org/10.3390/su11113128

Hurwitz, J. and Kirsch, D. (2018) *Machine Learning for Dummies*. Wiley. ISBN 1119245516, 9781119245513.

Huseynov, S., Ribera, L. A. and Palma, M. A. (2018). *Predicting the highway costs index with machine learning*. Available at: https://www.semanticscholar.org/paper/Predicting-the-Highway-Costs-Index-with-Machine-Huseynov-Ribera/8a278f7ebeea2d58a206302ec05246aa442f8293 (Access: 3 February 2020).

Ivanova, E. and Masarova, J. (2013) Importance of Road Infrastructure in the Economic Development and Competitiveness. *Economics and Management*, 18(2), 263-274. https://doi.org/10.5755/j01.em.18.2.4253

Kelleher, J. D., Namee, B. M. and D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. The MIT Press, 1st ed. ISBN-13: 978-0262029445.

Martin, L. D. (2015). *Survey of the use of pavement management systems in Brazil and case of study* (Levatamento da Utilização de Sistemas de Gerência de Pavimentos no Brasil e Estudo de Caso) (Dissertation), UFRJ/COPPE, Brazil, Program of Civil Engineering. Available at: http://www.coc.ufrj.br/pt/documents2/mestrado/2015/2595-martins-ld-tm-15-pdf (Accessed: 9 November 2019).

Mitchell, T. M., Carbonell, J. G. and Michalski, R. S. (2016). *Machine Learning*. Kluwer Academic Publishers. ISBN-13: 978-0934613095.

Mourougane, A. and Pisu, M. (2011). Promoting Infrastructure Development in Brazil, *OECD Economics Department Working Papers*, No. 898, 2011. OECD Publishing, Paris. https://doi.org/10.1787/5kg3krfnclr4-en

OECD. (2017). Transport Infrastructure, in Gaps and Governance Standarts of Public Infrastructure in Chile. *Infrastrucuture Governance Review*, OECD Publisching, Paris. https://doi.org/10.1787/9789264280700-en

Oladele, A. S., Adedmila, A. S. and Egwurube, J. A. (2011). Highway maintenance cost estimation modeling for developing countries: a case study of Nigeria. *Botswana Journal of Technology, October 2011*. Available at: https://www.researchgate.net/publication/301676255_HIGHWAY_MAINTENANCE_COST_ESTIMATION_MODELING_FOR_DEVELOPING_COUNTRIES_A_CASE_STUDY_OF_NIGERIA (Accessed: 20 December 2019).

Piryonisi, S. M. and El-Diraby, T. E. (2017). A data analytics solution for predicting the condition of roads using the most affordable attributes. *Canadian Society of Civil Engineering* (CSCE), Vancouver. Available at: https://www.researchgate.net/publication/318788153_A_data_analytics_solution_for_predicting_the_pavement_condition_index_of_roads_using_the_most_affordable_attributes (Accessed: 3 November 2019).

Pivatto, D., Fernandez, R. N., Saulo, H. and Carraro, A. (2017) Estimating the optimal time for a road concession contract in Brazil. *International Journal of Economics and Finance*, 9(12), 44-53. https://doi.org/10.5539/ijef.v9n12p44

Raschka, S. (2015) *Python Machine Learning*. Packt Publishing Ltd. ISBN 1783555149, 9781783555147.

Shalev-Shwartz, S. and Ben-David, S. (2014) *Understanding Machine Learning: from theory to algorithms*. Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

Sodikov, J. (2005). Cost estimation of highway projects in developing countries: artificial neural network approach. *Journal of the Estern Asia Society for Transportation Studies*, 6, 1036-1047. https://doi.org/10.11175/easts.6.1036

Wilmot, C. G and Mei, B. (2005). Neural Network modeling of highway construction costs. *Journal of Construction Engineering and Management*, 131(7). https://doi.org/10.1061/(ASCE)0733-9364(2005)131:7(765)

Wolff, M. G. C. and Caldas, M. A. F. (2018). A model for the evaluation of Brazilian road transport: a sustainable perspective. *Journal of Advanced Trasportation*, 2018, Article ID 5274789. https://doi.org/10.1155/2018/5274789