



**HAL**  
open science

# Environmental Impact Classification of Perishable Cargo Transport Using Data Mining

Manoel Eulálio Neto, Irenilza De Alencar Nääs, Nilsa Silva Lima

► **To cite this version:**

Manoel Eulálio Neto, Irenilza De Alencar Nääs, Nilsa Silva Lima. Environmental Impact Classification of Perishable Cargo Transport Using Data Mining. IFIP International Conference on Advances in Production Management Systems (APMS), Aug 2020, Novi Sad, Serbia. pp.624-630, 10.1007/978-3-030-57997-5\_72 . hal-03635656

**HAL Id: hal-03635656**

**<https://inria.hal.science/hal-03635656v1>**

Submitted on 20 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Environmental Impact Classification of Perishable Cargo Transport Using Data Mining

Manoel Eulálio Neto<sup>1,2</sup> [0000-0002-6047-0288], Irenilza de Alencar Nääs<sup>1</sup> [0000-0003-0663-9377] and Nilsa Duarte da Silva Lima<sup>1</sup> [0000-0002-1284-7810]

<sup>1</sup> Graduate Studies in Production Engineering, Paulista University,  
São Paulo, SP 04026-002, Brazil

<sup>2</sup> Centro Universitário Santo Agostinho,  
Teresina, PI 64.019-625, Brazil

manoeleulalio@gmail.com, irenilza@gmail.com,  
nilsa.lima@stricto.unip.br

**Abstract.** The study presents a model for classifying the environmental impact caused by the transport of vegetables from the production centers of several Brazilian states to a distribution center in Teresina, Brazil, using data mining. The distances from production regions to the distribution center were calculated. CO<sub>2</sub>-eq emissions and Global Warming Potential (GWP) were estimated. The GWP indicates the potential for the environmental impact that gas causes in each period (usually 100 years). We applied the data mining approach using the Rapid Miner Studio<sup>®</sup> software to build up the models. The target was the environmental impact indexed as “low,” “average,” and “high.” Results indicated that considering the on-road modal transport presented in the trees, the “product,” “distance,” and “quantity” classification for high environmental impact depends on the amount of product transported as well as the distance traveled. The found trees classify the impact and can be used as guidance for the decision-maker, as it can be used when planning and purchasing fruit and vegetables for public consumption.

**Keywords:** Classifiers, Random forest, transportation of vegetables, food supply chain.

## 1 Introduction

The Global Warming Potential (GWP) compares the amount of heat retained by a certain mass of CO<sub>2</sub> gas to the amount of heat captured by a similar mass of carbon dioxide along a period, usually 100 years. The GWP of an anisotropic gas indicates the potential for the environmental impact that gas causes in each given period. The report by the Intergovernmental Panel on Climate Change [1] estimates that greenhouse gas (GHG) emissions should be reduced by 40 to 70% by 2050 compared to 2010 levels, to avoid an increase of more than 2 °C at the global average temperature. Such an objective can avoid the more severe climate and, eventually, reduce the impacts. However, without a

clear research policy, initiatives to overcome the problems of large emerging economies and large territorial countries (for example, China, India, and Brazil) that depend on road transport is a significant challenge [2, 3].

The transport of fruit and vegetables in Brazil takes place from the producing region to regional distribution centers (Ceasa). Consumer centers are located close to large urban centers, usually in the capitals of Brazilian states. Transport is the logistical activity that plays a vital role within the supply chain. It enables sectors of the economy and the use of trucks to transport cargo can provide, in several situations, greater flexibility in the operations of distributing products and inputs, in a more agile way, because it can cover almost all regions of the country.

With the development of computers and automation, the storage and recovery of large volumes of data have been increased. As a result, machine learning techniques, including data mining, have become a useful tool for identifying and exploring patterns and relationships between many variables [4, 5]. Data mining applications offer classification models in some research areas, including health diagnosis and prognosis [6] and identifying gaps in education data [7].

The present study aimed to develop environmental impact estimation models using the data mining approach. The subject was the on-road cargo transportation, from part of the logistics chain for the distribution of fruit and vegetables (tomatoes, lettuce, bell peppers, cucumbers, and cabbage) from the production centers in several Brazilian states to the 'Nova Ceasa' distribution center in Teresina, Piauí State, Northeast Brazil.

## 2 Material and Methods

The products transported (tomatoes, lettuce, bell peppers, cabbage, and cucumbers) are produced in different regions of the country and distributed in Teresina by 'Nova Ceasa,' which is the fresh food distribution center. These products are transported in 'road modes' using cargo trucks from the farms in production centers to the distribution center from January to October 2019. The production distances were calculated, assuming that the transport trucks consume an average of 10 L/km of fuel (diesel). CO<sub>2</sub>-eq emissions and GWP were calculated in an online spreadsheet [8]. The values found for GWP were discretized into three conditions "low," "average," and "high." The data obtained were inserted and organized in an Excel<sup>®</sup> spreadsheet and further used in data processing using the data mining approach.

The Random forest approach in the data mining scenario was selected to describe the environmental impact of on-road transportation of fresh food perishable products. The Random forest classifier used for this study consists of using randomly selected features or a combination of features at each node to grow a classifying tree.

The complete dataset (containing the distances, km; amount transported, t; and the discretized values of GWP) was used to build a decision tree using Rapidminer<sup>®</sup> Studio, a software-based on Java version 9.2 (RapidMiner, Inc. Boston, Mass., USA). The alternative classification was found while processing the dataset, considering the final focus (target) as the environmental impact of cargo transport (GWP).

The operators used were 'retrieve data,' 'split data,' and 'Random forest' as a classifying algorithm. In the present study, we used 70% of the data to train the algorithm and 30% to develop the model as input values for the 'split' operator. The subsequent set of characteristics in the data training set is recognized by the attributes that discriminate several samples more precisely [9]. Amongst the trees found within the Random forest applied algorithm, three trees were selected, the first focusing on the transported product, the second on the transport distance, and the third focusing on the quantity of product transported.

The percentage of correctly classified samples compared to the number of all examples is named accuracy (Eq. 1).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

Figure 1 shows the data processing scheme of the present study.

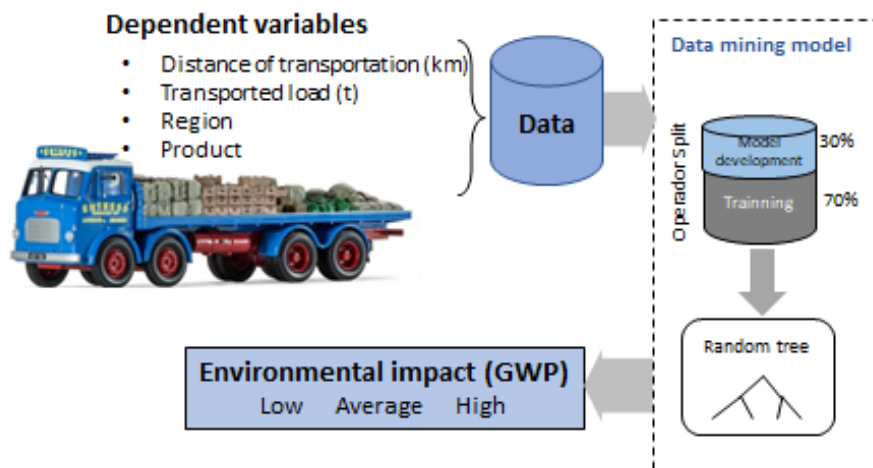
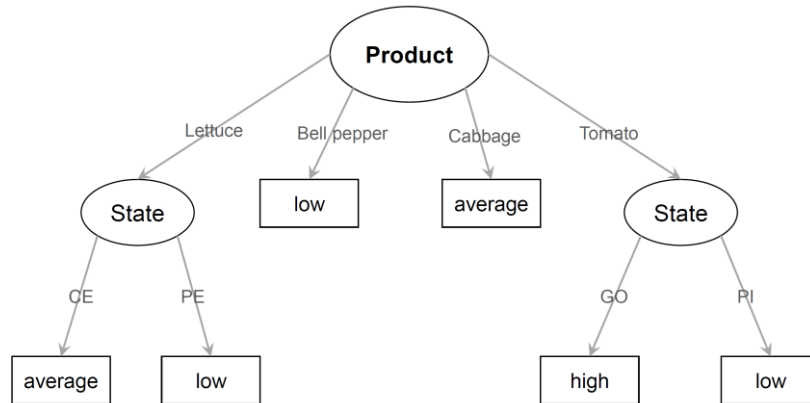


Fig. 1. Schematic of data processing used in the present study.

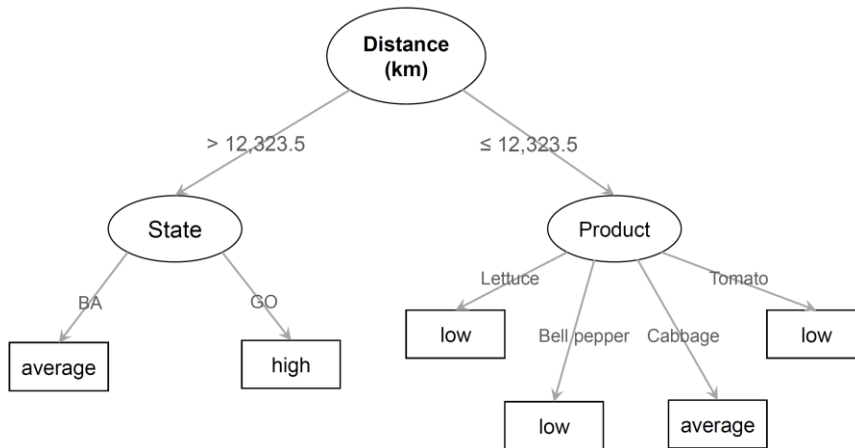
### 3 Results and Discussion

The trees found were selected for the model accuracy (75%), containing the classification of environmental impact (high, average, and low). The trees were selected whose processing classified the 'product,' 'distance,' and 'quantity' transported. Figures 2, 3, and 4 show the selected trees that were obtained from the processing results of the Random forest.



**Fig. 2.** Random tree with a focus on the environmental impact mainly due to the product transported.

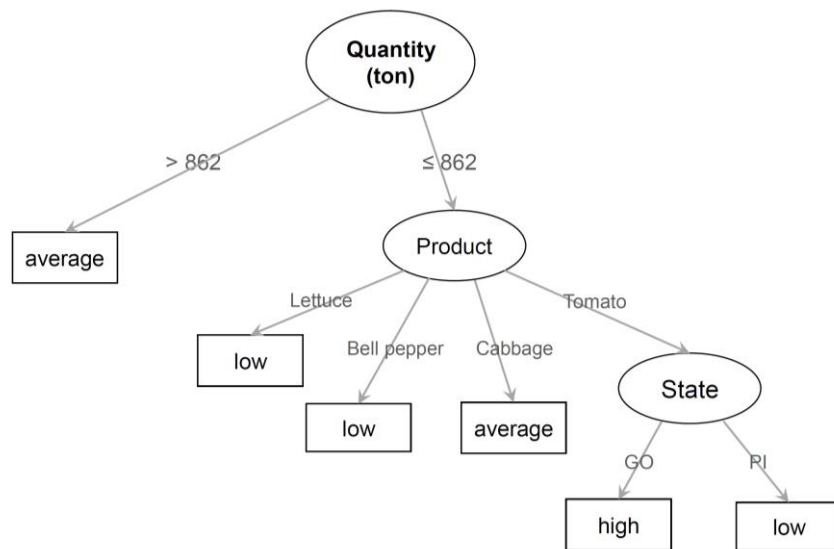
The classifying tree shown in Figure 2 indicates the level of impact on CO<sub>2</sub> emissions, considering the concepts, arbitrated, pre-established in this work (A = high, M = medium, and B = low). When analyzing the origin of the products by the state (or region) of production, it appears that for the tomato production from the State of Goiás (GO), it has a high environmental impact compared to the GWP of the other states that supply tomato products. It is highlighted the transport from Pernambuco (PE) with low impact.



**Fig. 3.** Random tree with a focus on the environmental impact, mainly due to the distance the products were transported.

The analysis of the tree concerning the distance between the producers and the distribution center resulted that for a distance less than or equal to 12,323.5 km, the lettuce, bell pepper, and cabbage products have a “low” environmental impact (GWP). In contrast, the tomato product produces a “low” impact. When the distance is more significant than 12,323.5 km, for the same products, it is seen that, when importing from the State of Goiás (GO), there is a “high” impact on CO<sub>2</sub> emissions. On the other hand, when acquired the product in the State of Bahia (BA), the environmental impact is “average.” Thus, the presented tree objectively guides the analyzed consequences of purchasing products from a particular region, giving decision-makers the possibility to make the choices that best protect the environment

The resulting tree (Fig. 4) focus the analysis of the impact concerning the quantity transported on-road from the production area to the distribution center in Teresina. When the tree is analyzed, it is noticed that for amounts less than or equal to 862 (t), the transport of the tomato product, originating in the State of Goiás (GO), has a high environmental impact. However, when provided in the State of Piauí (PI), the effect is classified, in the model, as low, clearly indicating the influence that the distance traveled has, since there is a higher consumption of fossil fuel, in the modal transportation used, with direct reflection in the GWP. It is also seen that for quantities less than or equal to 862 (t) for lettuce and bell peppers, the impact is low. On the other hand, when the amount exceeds 862 (t), the GWP presents a medium environmental impact, and the decision-maker might, given these indicators, make a choice that best protects the environment, not disregarding other variables considered by the consumers.



**Fig. 4.** Random tree with a focus on the environmental impact mainly due to the quantity of products transported.

The claims of ‘local food’ chains propose improvements in the relations between the urban and rural areas [10]. Within the ‘local food’ debate, environmental impacts are not restricted to GHG emissions. However, carbon reduction policies may provide potential trade-offs in the overall ecological sustainability issue until other impacts are measured [11].

Policy contexts need to address how environmental impact will be assessed and managed as governments attempt to meet the schedule of emissions cuts [12]. In a similar study, [13] provides a scenario of food on-road transportation in Brazil, stating that such an array of food distribution without impacting the environment is a difficult task in continental countries.

## 4 Conclusions

The objective of the present study was the development of an environmental impact estimation model (GWP) using data mining to transport loads in the road mode. This is part of the logistics chain for the distribution of fruit and vegetables from production centers in several Brazilian states to the ‘Nova Ceasa’ fresh food distribution center in Teresina, State of Piauí, Northeast of Brazil.

The models obtained as Random forest are useful to support tool for decision making since they can present graphical information to guide managers in their decisions. From the found trees, it was possible to verify that the vegetables from the Goiás state produce high environmental impact. The trees might guide the decision-maker, as they can be used when purchasing fruit and vegetables considering the environmental impact. It is reasonable to believe that the choice of the production center should favor the one that least contributes to the emission of polluting gases.

## References

1. IPCC- Intergovernmental Panel on Climate Change (2014). Fifth Assessment Report, p AR5
2. Muratori M, Smith SJ, Kyle P, Link R, Mignone BK, Kheshgi HS (2017) Role of the freight sector in future climate change mitigation scenarios. *Environ Sci Technol* 51:3526–3533.
3. Dente SMR, Tavasszy L (2018) Policy-oriented emission factors for road freight transport. *Transp Res D* 61:33–41.
4. Díez J, Albertí P, Ripoll G, Lahoz F, Fernández I, Olleta JL, Panea B, Sañudo C, Bahamonde A, Goyache F (2006) Using machine learning procedures to ascertain the influence of beef carcass profiles on carcass conformation scores. *Meat Sci* 73: 09-115.
5. Pereira DF, Miyamoto BCB, Maia GDN, Sales T, Magalhães MM, Gates RS (2013) Machine vision to identify broiler breeder behavior. *Comput Electron Agr* 99: 194-199
6. Zhuang X, Bi M, Guo J, Wu S, Zhang T (2018) Development of an early warning algorithm to detect sick broilers. *Comput Electron Agr* 144: 102–113.
7. Baker RSJ, Isotani S, Carvalho AMJB. (2011) Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19 (2).
8. CFC-Carbon Footprint Calculator (2018) Vehicle CO<sub>2</sub> emissions footprint calculator. <https://www.commercialfleet.org/tools/van/carbon-footprint-calculator>

9. Lavrac N. (1999). Selected techniques for data mining in medicine. *Artif. Intell. Med.*, 16: 3-23.
10. DuPuis M, Goodman D (2005) Should we go 'home' to eat? Towards a reflexive politics in localism. *J Rural Stud* 21: 359–371
11. Hall G, Rothwell A, Grant T, Isaacs B, Ford L, Dixon J, Kirk M, Friel S (2014) Potential environmental and population health impacts of local urban food systems under climate change: a life cycle analysis case study of lettuce and chicken. *Agr Food Secur* 3: 6
12. Rothwell A, Ridoutt B, Page G, Bellotti W (2016) Environmental performance of local food: trade-offs and implications for climate resilience in a developed city. *J Clean Prod* 114: 420-430.
13. Duarte GT, Nääs IA, Innocencio CM, Cordeiro AFS, Silva, RBTR (2019). Environmental impact of the on-road transportation distance and product volume from farm to a fresh food distribution center: a case study in Brazil. *Environ Sci Pollut Res* 26: 33694–33701.