



DYNA
ISSN: 0012-7353
ISSN: 2346-2183
Universidad Nacional de Colombia

A review of Machine Learning (ML) algorithms used for modeling travel mode choice

Pineda-Jaramillo, Juan D

A review of Machine Learning (ML) algorithms used for modeling travel mode choice

DYNA, vol. 86, no. 211, 2019

Universidad Nacional de Colombia

Available in: <http://www.redalyc.org/articulo.oa?id=49663345004>

DOI: 10.15446/dyna.v86n211.79743

A review of Machine Learning (ML) algorithms used for modeling travel mode choice

Una revisión de los algoritmos de Machine Learning (ML) utilizados para la modelación de la elección de modo de viaje

Juan D Pineda-Jaramillo a jd-pineda@unal.edu.co
Universidad Nacional de Colombia, Colombia

Abstract: In recent decades, transportation planning researchers have used diverse types of machine learning (ML) algorithms to research a wide range of topics. This review paper starts with a brief explanation of some ML algorithms commonly used for transportation research, specifically Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM) and Cluster Analysis (CA). Then, these different methodologies used by researchers for modeling travel mode choice are collected and compared with the Multinomial Logit Model (MNL) which is the most commonly-used discrete choice model. Finally, the characterization of ML algorithms is discussed and Random Forest (RF), a variant of Decision Tree algorithms, is presented as the best methodology for modeling travel mode choice.

Keywords: modeling travel mode choice, Artificial Neural Networks (ANN), Decision Trees (DT), Support-Vector Machines (SVM), Cluster Analysis (CA), Multinomial Logit Model (MNL), Machine Learning (ML) algorithms.

Resumen: En décadas recientes, los investigadores de planificación de transporte han usado diversos tipos de algoritmos de Machine Learning (ML, por sus siglas en inglés) para investigar un amplio rango de temas. Este artículo de revisión inicia con una breve explicación de algunos algoritmos de Machine Learning comúnmente utilizados para la investigación en transporte, específicamente Redes Neuronales Artificiales (ANN), Árboles de Decisión (DT), Máquinas de Vector de Soporte (SVM) y Análisis de Grupos (CA). Luego, estas diferentes metodologías usadas por investigadores para modelar la elección de modo de viaje son recogidos y comparados con el Modelo Logit Multinomial (MNL) el cual es el modelo de elección discreta más comúnmente utilizado. Finalmente, la caracterización de los algoritmos de ML es discutida y el Bosque Aleatorio (RF), una variante de los algoritmos de Árboles de Decisión, es presentado como la mejor metodología para modelar la elección de modo de viaje.

Palabras clave: modelación de la elección de modo de viaje, Redes Neuronales Artificiales (ANN), Árboles de Decisión (DT), Máquinas de Vector de Soporte (SVM), Análisis de Grupos (CA), Modelo Logit Multinomial (MNL), algoritmos de Machine Learning (ML).

1. Introduction

The transportation planning sector needs to model travel mode choice to predict travel demand and understand the causal variables [1]. Currently, the literature shows evidence that travel mode choice depends on a large number of variables including individual, household and exogenous factors like security and comfort on a trip, weather conditions and built environment [2-9].

DYNA, vol. 86, no. 211, 2019

Universidad Nacional de Colombia

Received: 17 May 2019

Revised document received: 20 August 2019

Accepted: 05 September 2019

DOI: 10.15446/dyna.v86n211.79743

CC BY-NC-ND

Different models of travel mode choice, such as discrete choice models, where the travel modes signify mutually exclusive and joint alternatives, have been used within diverse frameworks [10]. The multinomial logit model (MNL) is the most commonly-used discrete choice model for modeling travel mode choice [11,12]. It considers the principle of utility maximization and has a singular mathematical framework which allows for parameter estimation, and so it has been widely adopted in transportation planning [6,13].

However, the MNL model has several limitations because it assumes that the probability of each alternative is independent of the features of the rest of the alternatives [13,14].

Machine learning (ML) algorithms have been demonstrated to work well for statistical approaches used to model travel mode choice. ML algorithms do not make drastic assumptions about the studied data, but learn to represent non-linear and, in general, complex relationships in a data-driven way [15].

The usefulness of ML algorithms has been demonstrated for many different fields, including transportation planning. In this field, ML algorithms have been used for classifying accidents and studying safety and human behavior, among other applications [16-23].

This paper is organized thusly: First, it presents different machine learning algorithms used in transportation research. Section 2 outlines the most common machine learning algorithms used for transportation research. Section 3 details the most common discrete choice model used for modeling travel mode choice: the Multinomial Logit Model. Section 4 presents a comprehensive comparison of different ML algorithms used for modeling travel mode choice. Section 5 puts forward a discussion and the notable conclusions. Finally, Section 6 lists the references used to construct this review paper.

2. Machine learning (ML) algorithms for transportation research

The expression Machine learning (ML) is used to define a group of methods or algorithms that allow computers to mechanize data-driven model programming and build models by means of a methodical detection of patterns in statistically significant data [24]. In the 1930s, Thomas Ross made the first attempt to develop a machine that simulated the behavior of a living being [25]. Later, Samuel (1959) defined ML as a “field of study that gives computers the ability to learn without being explicitly programmed” [26].

The application of different ML techniques in the field of transportation intends to meet the challenges of growing travel demands, safety concerns, energy consumption, emissions, and environmental degradation [27].

ML algorithms can be classified as follows [28]:

- Supervised learning, where ML algorithms generate a function that charts input data to target output data.
- Unsupervised learning, where there is no target output data and the ML algorithm simply models a set of input data, looking for clustering in that data [29].
- Semi-supervised learning is a combination of both of the above, where ML algorithms use labeled data and unlabeled data.
- Reinforcement learning, where ML algorithms learn through their interaction in an environment. The ML algorithm obtains feedback about the accuracy of its response.
- Inductive learning is when the ML algorithms learn, based on previous knowledge, their own inductive bias.

In order to build an optimal predictive model, it is vital to consider the following specific phases [27]:

i. Design and data ingestion

This phase includes three steps: data preparation, exploratory analysis and feature extraction. In the first phase, data sources have to be assessed and used to incorporate the advance model into the problem to be solved. This phase is essential because it entails the initial evaluation of the data.

ii. Proof of concept

This second phase includes two steps: modeling algorithms and model evaluation. In this phase it is necessary to choose the model that best fits the data. It is important to take into account the basic differences of every model to achieve a good validation.

iii. Integrate and scale

This final phase includes two steps: initial pilot and full-scale implementation.

This phase represents real-time prediction information about the model's performance. In this last phase, the model must be continuously updated and adjusted to obtain the best results.

Below, the most popular Machine Learning techniques used in transportation research will be presented. However, it is important to note that there are other algorithms that are not so popular in this field that are not included here.

2.1. Artificial neural networks (ANN)

ANNs are used to extract complex patterns from the data, and perceive trends that are too complex to be observed by humans or other computer methods with their outstanding ability to derive meaning from data that is complex or inaccurate [30-32]. McCulloch and Pitts (1943) introduced the concept of ANN, and it was designed to simulate the functions and structure of the nervous systems in living beings [33].

ANNs are very powerful tools that have been used for numerous applications such as medicine [34-37], transportation [23,38-41], optimization [31,42-45], and even quantum physics [46-50] among others.

ANNs are composed of a large number of *neurons*, elements that are interconnected in parallel and work in unison to solve diverse problems.

The ANN is trained using *input* and *target* data. This process means that available target data is compared with *output* data provided by the ANN, and later, the ANN's parameters are adjusted by means of an iterative process until an optimal agreement between reality and the model is accomplished [31].

An example of the structure of an ANN is presented in Fig. 1, where the hidden layer (the first), has a determined number of neurons that need to be defined.

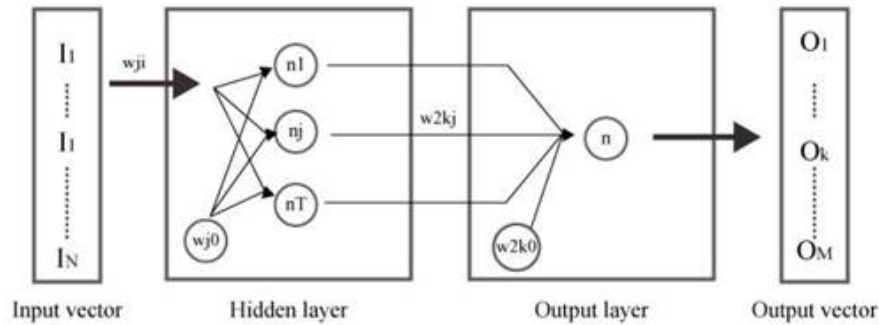


Figure 1.

Artificial neural network framework.

Source: The Author

The *output* layer (the second) has one neuron that is defined with a linear transfer function.

Eq. (1) presents the formulation of an ANN:

$$O_k = \tilde{g} \left(\sum_{j=0}^M w_{2kj} \cdot g \left(\sum_{i=0}^N w_{ji} \cdot I_i \right) \right) \quad (1)$$

In Eq. (1):

- O_k : ANN output.
- M : quantity of “output” elements.
- I_i : “input” data.
- N : quantity of input attributes.
- w_{ji} : synaptic weight of the first layer.
- w_{2kj} : synaptic weight of the second layer.

Synaptic weight w_{ji} describes the strength of a synaptic connection between the postsynaptic neuron i and the presynaptic neuron j . This structure is capable of recognizing non-linear relationships between *input* data and *output* data [32].

Other advantages of ANNs include [32,51,52]:

- Its adaptive learning: The skill to use the data given during training to learn how to do specific tasks.
- Its self-organization ability: A big advantage of ANNs is that they arrange their own information obtained during the learning process.
- Its real time operation: Another advantage is that computations of ANNs can be accomplished in parallel and special hardware devices have been designed which take advantage of this special skill.
- Its failure tolerance by means of redundant information coding: fractional destruction of an ANN leads to degradation of the performance.

There are different types of ANN, the most well-known are:

2.1.1. Self-Organizing Maps (SOM)

Self-Organizing Maps (SOM) are unsupervised ANN that reduce the input dimensionality with the aim of representing distribution as a “map”, where similar points are mapped carefully together [53].

SOMs are very useful for visualization due to the way they take high-dimensional data and create low-dimensional images of it.

It is possible to perform cluster analysis on the map itself, because it has thousands of nodes.

2.1.2. Convolutional Neural Networks (CNN)

CNN are commonly used for image processing applications. Energy, computational mechanics, electronic systems and remote sensing use CNN for analysis and prediction among other applications.

In other words, CNNs are regularized versions of fully connected networks, which means that each neuron in one layer is connected to all neurons in the next layer [54].

2.1.3. Recurrent Neural Networks (RNN)

RNN are Artificial Neural Networks that contain cyclic connections, making them a more powerful tool for modeling sequence data than standard ANN. This technique has proved to be an outstanding success in sequence labeling, speech recognition, language modeling and handwriting recognition [55].

It is important to briefly mention that the ANN algorithms detailed above are related to Deep Learning. Deep Learning is a term introduced by Ian Goodfellow et al. in 2016 [56], for “a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts, because the computer gathers knowledge from experience, there is no need for a human computer operator

to formally specify all the knowledge that the computer needs. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones; a graph of these hierarchies would be many layers deep.”

Deep learning algorithms have recently emerged from ML techniques, and these methods exploit much deeper and more complex architectures than ML algorithms, and can achieve better results than traditional methods in many fields [57].

2.2. Decision Trees

Decision trees (DTs) are oriented graphs formed by a finite number of nodes departing from the root nodes [58].

Decision trees are non-parametric methods with a similar structure to a flowchart or to a tree and they can be used to classify problems [24].

Decision trees are powerful algorithms for classifying data, where a tree structure is used for modeling the different relationships between the features and potential output data. This ML algorithm is so-called because it simulates a real tree, which begins at a wide trunk, and as it rises is divided into narrower branches. Similarly, DT use an architecture of branching choices, beginning with the main question for a specific problem which needs to be answered to solve that problem, later a secondary question must to be answered to continue disaggregating the data and classifying outcomes.

For a better understanding of how a DT works, consider the decision tree presented in Fig. 2, which predicts if it is a good idea to make a trip by bike from an origin to a destination. The bike trip to be considered starts at the origin node, where it is then passed through different decision nodes that need to be assessed, considering the characteristics of external conditions (rain, travel time or topography). These assessments divide the data into different branches that indicate probable outcomes of a decision, represented as Yes or No results. It is possible for these decisions to have more than two alternatives.

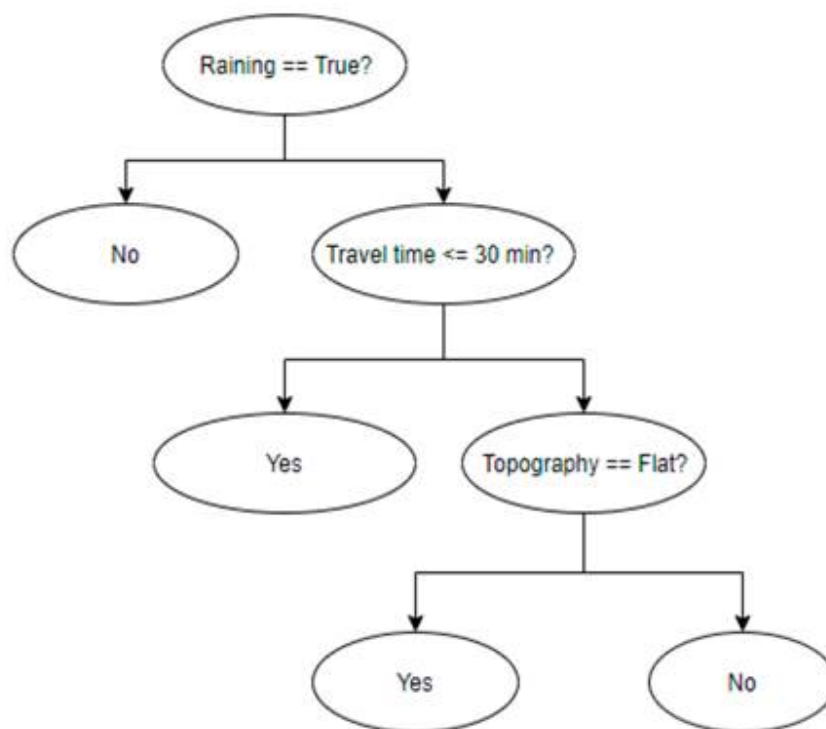


Figure 2.
Decision tree example.
Source: The Author.

When a final decision is made, the DT ends with leaf (or terminal) nodes denoting the action to be taken as the final result of the series of choices.

A great advantage of DT is that the flowchart-like tree structure is not exclusively for internal use. When a DT algorithm is built, it is easily interpreted by those without technical knowledge.

This feature means that a model can be assessed as to whether it works well enough for a specific task.

In addition to this, DTs can be interpreted easily, and they can deal with non-linear relationships and interactions between every variable. However, DTs are very sensitive to noisy data, and also tend to overfit the data, rendering it useless [59]. Tree-based algorithms combine various DTs in order to build more accurate and steady classifiers than simple DTs [60].

To summarize, some advantages of DTs are:

- Their functioning is easy to understand and interpret.
- They require little data preparation from the user to build an optimal DT. There is no need to apply normalization to the data.

On the other hand, the main disadvantage of DT is that they have a great probability of overfitting noisy and defective data, and this probability increases as the tree gets deeper and more complex.

Some potential uses of DT algorithms include:

- Diagnosis of medical conditions based on symptoms or laboratory measurements in medicine.
- Credit scoring models for banking agencies.
- Marketing studies of customer behavior for advertising agencies.
- Modeling travel mode choice, as will be seen later in this paper.

In general, DT algorithms are one of the most-used ML techniques, and they can be applied to model many types of data [61].

Bagging is a straight-forward tree-based algorithm method, whence several DTs are trained at the same time in parallel using bootstrap samples of the data. Class assignments for the final prediction are determined by the majority vote of all trees running in parallel [62].

Random forests (RFs) are other tree-based algorithms that are associated with bagging. While RFs train several decision trees in parallel using bootstrap samples of the data, each split at the nodes of the trees is calculated by a random subdivision of features. Like bagging, RFs determine class assignment for predictions through the majority vote of the ensemble of trees [63].

2.3. Support-Vector Machines

Support Vector Machines (SVMs) are another ML algorithm method used for binary classification. SVMs can be defined as risk-based supervised learning algorithms for classifying data patterns by identifying a frontier with a maximum margin within data of the same class [24,64].

SVMs are considered to be supervised learning algorithms; when they are given labeled training data, the SVM outputs an optimum flat boundary called a *hyperplane*. The hyperplane is simply a line splitting a plane into two portions in a two-dimensional space, where each class lies either side of the line.

An SVM algorithm classifies data by projecting the target variables into a high-dimensional feature space, where classes are linearly separable [65]. It is possible to imagine an SVM as a surface that generates a limit between datapoints plotted multidimensionally, representing samples and their respective attribute values.

An SVM, in order to perform linear classification, overcomes a non-linear classification, indirectly plotting its inputs in high-dimensional feature spaces.

If data is unlabeled, is not possible to use a supervised learning algorithm, and an unsupervised learning method is required, which tries to discover natural groupings of data to assemble and then plot new data to the ensembles formed. There are SVM algorithms that use the statistics of Support Vectors to classify unlabeled data. They are called Support Vector Clustering [66].

Support-Vector Machines can be adapted for use with almost any type of learning task, including prediction and classification. Many of the algorithm's key successes have been in pattern recognition of data. Prominent applications include [61]:

- Text categorization to identify the language used in texts.
- Detection of events like earthquakes or security breaches.
- Discovery of uncommon and important events, like combustion engine failure.
- Classification of microarray gene expression data, for identifying cancer and other important diseases.
- Classification of texts by subject.
- Modeling travel mode choice, as will be seen later in this paper.

2.4. Cluster Analysis

Cluster Analysis (CA) is an unsupervised machine learning technique used to divide the data into similar groups with similar features, with the aim of maximizing the heterogeneity between clusters (groups) and the similarities between in-cluster samples [67,68].

It divides data into separated clusters without first having been told how the clusters should look. As it is an unsupervised ML algorithm, CA is issued for knowledge detection rather than prediction. It offers an insight into the natural grouping of the data.

Latent Class Clustering (LCC) is a particular method with advantages over regular CA, similar to Ward's method and k-means. These advantages include access to much statistical criteria used for deciding the suitable number of clusters, and the ability to use different types of features with no need for previous standardization which could modify the outcomes [69].

The relevance of CA lies in that the clusters can then be used for action. For example, CA are employed to [61]:

- Perceive anomalous behavior, such as unauthorized network intrusions, by recognizing different patterns of use that fall outside the known groups.
- Divide customers into clusters with similar socioeconomic aspects or buying patterns for advertising campaigns.
- Simplify large datasets by clustering features with similar values into a smaller number of homogeneous categories.

CA is useful whenever differences in the data can be exemplified by a small number of clusters. CA reduces complexity and give insight into patterns of relationships.

The k-means clustering algorithm is the most popular CA algorithm and serves as the foundation for many sophisticated clustering techniques. It is popular because it uses simple principles which can be described without using statistical concepts. K-means is highly flexible and can be modified using simple changes to overcome all of its shortcomings and so achieve optimal results in several real-world cases.

On the other hand, the main weaknesses of k-means lie in that it is not as sophisticated as modern cluster techniques because it uses a component of random chance, and that it will find the best set of groups is not

guaranteed. The other disadvantage is that it relies on estimation to assign the number of groups for the data [61].

CA algorithms (including k-means) have been used in different fields of transportation engineering with optimal outcomes [70-74]. Some authors [7,75] used LCC analysis to segment heterogeneous traffic accident data sets into homogenous accident. De Oña et al. [58] used a CA method to assess passenger heterogeneity, where the CA method stratified the sample of passengers into clusters with similar features and therefore into clusters of homogeneous perceptions concerning the service. Other authors [76] used a CA to analyze the effect of workplace relocation on an individual's travel behavior and activity.

3. Multinomial Logit Models for modeling travel mode choice

Multinomial Logit Models (MNL), and a large number of variations on them, are extensively used for modeling travel mode choice [11,14,77-79].

The existence of an individual n and a set of m variables $X_n = \{X_{1n}, \dots, X_{mn}\}$ can define a choice set C_n of I alternatives and corresponding utility functions (Eq. 2 and Eq. 3):

$$U_{in} = \beta_1 \cdot X_{1in} + \dots + \beta_m \cdot X_{min} + \varepsilon_{in} \quad \forall i \in C_n \quad (2)$$

$$U_{in} = V_{in} + \varepsilon_{in} \quad (3)$$

Where:

- $\varepsilon_{1n}, \dots, \varepsilon_{mn}$ are independent and identically distributed random variables (iid). In other words, if all variables have the same probability distribution, and every variable is mutually independent of each other, it is said that the sequence of variables is iid.
- β_1, \dots, β_m is the set of parameters to be estimated. This is carried out by means of a minimization of the negative log-likelihood (the logarithm of the likelihood function, the function that estimates a parameter from a set of statistics) (Eq. 4):

$$L = - \sum_{n=1}^N \sum_{i \in C_n} y_{in} \log [P(i|C_n)] \quad (4)$$

- where $y_{in} = 1$ if individual n chooses i . Otherwise, $y_{in} = 0$. The probability of choosing $i \in C_n$ for MNL is presented in Eq. (5):

$$P(i|C_n) = P(U_{in} > \max_{j \neq i} (U_{jn})) = \frac{\exp^{V_{in}}}{\sum_{j \in C_n} \exp^{V_{jn}}} \quad (5)$$

This mathematical model has deep theoretical foundations [10] making widespread use of ε to express statistical properties.

One important disadvantage of these type of models compared to ML techniques is that a Logit model typically focuses on parameter estimation and does not lend enough importance to prediction. On the other hand, one big advantage of Logit models compared to ML algorithms lies in that ML algorithms are built for predicting values, but are frequently considered to be difficult to infer and are almost never used to extract behavioral findings from the model outputs [80].

4. Machine Learning (ML) algorithms for modeling travel mode choice

Several ML algorithms have been used for modeling travel mode choice in recent decades. This paper will cover the most important ones.

Regarding ANNs, Shmueli et al. (1996) [81] compared a simple Multilayer Perceptron (MLP) to non-linear classification and regression trees. After this comparison they demonstrated that both methodologies perform similarly, and they perform optimally when modeling travel mode choice.

Later, Sayed and Razavi (2000) [82] used fuzzy artificial neural networks, demonstrating that they can be used to classify in the same way that Probit and Logit models are usually used to model travel mode choice.

Mohammadian and Miller (2002) [83] compared the performance of MLP and Nested logit models, showing that the first has a significant advantage over the second when the percentage of properly classified instances for predicting domestic vehicle choices are considered.

Vythoulkas et al. (2003) [84] showed that outcomes from fuzzy artificial neural networks that model travel mode choice compare positively to a Logit model.

Cantarella and De Luca (2003) [38] trained two ANNs with different frameworks to model travel mode choice. They showed that both artificial neural networks perform better than a Multinomial Logit model.

Other authors, like Hensher and Ton (2000) [85], Xie et al. (2003) [86], Andrade et al. (2006) [87], Celikoglu (2006) [39] and Zhang and Xie (2008) [88] demonstrated that the predictive capability of MLP is superior to multinomial and nested logit models, concluding that MLP could overcome the utility function estimation in the modeling of travel mode choice.

Zhao et al. (2010) [40] have shown that the precision of probabilistic artificial neural networks is comparable to basic artificial neural networks for predicting travel mode choice, while Omrani et al. (2013) [41] posited that ANNs are more precise than the other alternatives that they examined.

Pulugurta et al. (2013) [89] found that fuzzy ANN were better for detecting and including human knowledge and cognitive activities into mode choice behavior.

On the other hand, there are some publications that come to the opposite conclusion, like Abdelwahab and Abdel-Aty (2002) [20] who proved that a two-level nested logit model beats the Multi-Layer Perceptron (ANN) in the analysis of driver injury severity, or Teng and Qi (2003) [90, 91] who presented the supremacy of wavelets over diverse artificial neural network frameworks for modeling incident detection.

Decision trees (DTs) have also been applied for modeling travel mode choice. For example, Xie et al. (2003) [86] compared DTs and ANNs with MNL models, concluding that DTs and ANNs outperform MNL models. Additionally, they found that DTs are more effective and can be better interpreted than Artificial Neural Networks.

On the other hand, Rasouli et al. (2014) [92] explored the connection between predictive performance and the number of decision trees by means of ensemble learning. Results of this study suggest that the accuracy of predicting transport mode choice is improved, although non-monotonically, with increasing ensemble size.

Tang et al. (2015) [93] used DTs to explore travel mode choice in cases in which individuals have only two options to choose between, looking to understand people's mode-switching behavior. In this paper it was demonstrated that DT outperforms MNL models in predictive capability.

Zhan et al. (2016) [94] used hierarchical tree-based models for exploring the travel characteristics of students from China, determining which variables influenced their travel mode choice.

Ravi Sekhar et al. (2016) [95] used a random forest DT to model travel mode choice in Delhi, by means of 5000 stratified household samples collected in the city using household interview surveys.

Hagenauer and Helbich (2017) [62] proved that a decision tree framework, specifically a random forest algorithm, outperforms any of the other classifiers that were investigated, even an MNL model.

Cheng et al. (2019) [96] applied a random forest algorithm to model travel mode choice behavior, obtaining outstanding results.

SVM algorithms have also been used to model travel mode choice, for instance, Zhang and Xie (2008) [88] compared Support-Vector Machines, Artificial Neural Networks and Multinomial Logit models for modeling travel mode choice and demonstrated that Support-Vector Machines provided the highest accuracy of every tested model. On the contrary, Omrani (2015) [97] demonstrated that Artificial Neural Networks have a higher accuracy than Support-Vector Machines and Multinomial Logit models for predicting the travel mode of individuals in the city of Luxembourg.

Additionally, Xian-Yu (2011) [98] demonstrated that an SVM model has fast convergence and high precision, outperforming ANN and nested logit models, which is a very important consideration for modeling travel mode choice.

Regarding CA, Ding and Zhang (2016) [99] estimated travel behaviors by dividing individual travelers into several groups based on their personal features using CA.

Li et al. (2016) [100] implemented a cluster-based logistic-regression model to predict travel mode choice during holidays in Beijing, where they employed a regression and a DT method to split the source data into groups, and they concluded that since the cluster-based logistic regression model evades the variable interaction effects, it outperforms the logistic-regression model in its prediction accuracy.

Also, in 2016, Pirra and Diana [101] used CA to socioeconomically characterize different profiles of travelers in the U.S. with specific kinds of tours.

On the other hand, Molin et al. (2016) [102] conducted a latent class cluster analysis to identify multimodal travel groups based on the self-reported incidence of mode use, finding that most car drivers have negative attitudes towards bicycles and public transport, while car drivers who often use public transport have more positive attitudes to bicycles and public transport.

The above-mentioned publications are exhaustive contributions to applications of ML methodologies for modeling travel mode choice, however, although there are a huge number of Machine Learning classifiers available, these investigations only deal with a limited set of these [103]. Along these lines, Hagenauer and Helbich (2017) [62] carried out a comparative study of seven ML classifiers for modeling travel mode choice (included the commonly used MNL), showing that an advanced classifier like the random forest significantly outperforms all the other investigated classifiers. Regarding this assumption, Fernández-Delgado et al. (2014) [103] have proven that random forests can produce highly accurate outcomes for many applications.

Table 1 presents different approaches used by the authors mentioned here to model travel mode choice using Machine Learning algorithms.

Table 1.
Analysis of literature on modeling travel mode choice.

References	Problem ¹	Architecture ²	Training ³
Cheng et al. (preprint 2019) [96]	CLA	RF	Other
Hagenauer et al. (2017) [62]	CLA	RF	Other
Zhan et al. (2016) [94]	CLA	Hybrid	Other
Molin et al. (2016) [102]	CLA	Hybrid	Other
Li et al. (2016) [100]	CLA	Hybrid	Other
Ding and Zhang (2016) [99]	CLA	Hybrid	Other
Pirra and Diana (2016) [101]	CLA	Hybrid	Other
Ravi Sekhar et al. (2016) [95]	CLA	RF	Other
Tang, et al. (2015) [93]	CLA	Hybrid	Other
Omrani (2015) [97]	FA	SVM	BP
Rasouli and Timmermans (2014) [92]	CLA	RBF	LM
Pulugurta et al. (2013) [89]	FA	F	Other
Xian-Yu (2011) [98]	FA	SVM	BP
Dia and Panwai (2010) [104]	FA	F	BP
Zhang and Xie (2008) [88]	FA	SVM	BP
Dia and Panwai (2007) [105]	CLA	Hybrid	Other
Andrade et al. (2006) [87]	FA	F	Other
Celikoglu (2006) [39]	CLA	RBF	LM
Xie et al. (2003) [86]	CLA	MLP	BP
Cantarella and de Luca (2003) [38]	CLA	MLP	BP
Vythoulkas and Koutsopoulos (2003) [84]	FA	F	BP
Mohammadian and Miller (2002) [83]	CLA	MLP	BP
Nijkamp et al. (1996) [106]	FA	MLP	BP
Shmueli et al. (1996) [81]	CLA	MLP	BP

1: CLA: Classification, FA: Function Approximation.

2: RF: Random Forests, SVM: Support Vector Machines, F: Fuzzy, RBF: Radial Basis Function, MLP: Multi-Layer Perceptron.

3: BP: Back-Propagation, LM: Levenberg-Marquardt.

Source: The Author.

5. Discussion and conclusions

With the increasing popularity of ML algorithms in transportation research, there are many questions related to their advantages and disadvantages when compared to the Logit models that are commonly used to model travel mode choice. For this reason, this paper presents a comparison of these methodologies.

When compared to different ML algorithms that model travel mode choice, ANNs, DTs, SVMs and CA algorithms perform exceptionally well, better than MNL for almost every case. In addition, if multiple Decision Trees are combined in a Random Forest algorithm, its outcomes are better than any other machine learning algorithm.

The better performance of Random Forest can be attributed to its flexibility and power combined into a single ML method. As ensemble use is just a minor, random portion of the full feature set, RF approaches can handle enormous datasets, where other models might fail. At the same time, error rates for the majority of learning tasks are similar to any other approach.

ANN models have been applied in countless research projects for modeling travel choice mode, but in many cases, researchers implement this algorithm blindly, disregarding some of its inadequacies such as its intrinsic inability to present an exclusive solution to an issue (for this reason, it is common for many researchers to refer to ANN models as “black-box models”) [107].

For many authors, the lowest accuracy was given by the MNL model, demonstrating its less effective modeling abilities for modeling travel mode choice.

There is great potential in the integration of ideas from Logit and ML algorithms (and the exploration of ideas from Deep Learning techniques as well) to develop sophisticated models for modeling travel mode choice and proposing some possible research avenues, like, for example:

- Exploring which ML technique is most suitable for modeling travel mode choice.
- Incorporating the behavioral assumption that considers alternative-specific, which is enabled by the data framework of logit models “layered” into ML techniques.
- Exploring Deep Learning techniques for modeling travel mode choice.

References

- [1] Ortúzar, J. and Willumsen, L., *Modelling Transport* Chichester., John Wiley and Sons, 2011.
- [2] Ben-Akiva, M., Walker, J . Bernardino, A ., Gopinath, D ., Morikawa, T. and Polydoropoulou A ., *Integration of choice and latent variable models in Perpetual Motion: Travel behaviour research opportunities and challenges*, Amsterdam, 2002.
- [3] Dieleman, F., Dijst, M. and Burghouwt, G., Urban form and travel behaviour: micro-level household attributes and residential context. *Urban Studies*, 39(3), pp. 507-527, 2002. DOI: 10.1080/00420980220112801
- [4] Schwanen, T. and Mokhtarian, P., What affects commute mode choice: Neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*, 13(1), pp. 83-99, 2005. DOI: 10.1016/j.jtrangeo.2004.11.001
- [5] Böcker, L., Van Amen, P. and Helbich, M., Elderly travel frequencies and transport mode choices in greater Rotterdam, the Netherlands. *Transportation*, 44(4) pp. 831-852, 2016. DOI: 10.1007/s11116-016-9680-z
- [6] Böcker, L., Dijst, M. and Prillwitz, J., Impact of everyday weather on individual daily travel behaviours in perspective. a literature review. *Transport Reviews*, 33(1), pp. 71-91, 2013. DOI: 10.1080/01441647.2012.747114
- [7] Arbeláez, O., *Modelación de la elección de la bicicleta pública y privada en ciudades*, MSc. Thesis, Department of Civil Engineering, Universidad Nacional de Colombia, Medellín, 2015.
- [8] Ewing, R. and Cervero, R., Travel and the built environment: a meta-analysis. *Journal of the American. Planning Association*, 76(3), pp. 265-294, 2010. DOI: 10.1080/01944361003766766
- [9] Sprumont, F., Viti, F., Caruso, G. and König, A., Workplace relocation and mobility changes in a transnational metropolitan area: the case of the University of Luxembourg. *Transportation Research Procedia*, 4, pp. 286-299, 2014. DOI: 10.1016/j.trpro.2014.11.022
- [10] Ben-Akiva, M. and Lerman, S., *Discrete choice analysis: theory and application to travel demand*. MIT Press, Boston, 1985.
- [11] Pineda-Jaramillo, J.D., Sarmiento, I. and Córdoba, J.E., Railway and road discrete choice model for foreign trade freight between Antioquia and the Port of Cartagena. *Ingeniería e Investigación*, 36(3), pp. 22-28, 2016. DOI: 10.15446/ing.investig.v36n3.57370
- [12] Rich, J., Holmblad, P. and Hansen, C., A weighted logit freight mode-choice model. *Journal of Transportation Research Part A: Policy and Practice*, 45(6), pp. 1006-1019, 2009. DOI: 10.1016/j.jtre.2009.02.001
- [13] Ortúzar, J. y Román, C., El problema de modelación de demanda desde una perspectiva desagregada: el caso del transporte. *Eure*, 29(88), pp. 149-171, 2003. DOI: 10.4067/S0250-71612003008800007
- [14] McFadden, D., Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics*, New York, Academic Press, 1973, pp. 105-142.

- [15] Bishop, C., Pattern recognition and Machine Learning, New York, Springer, 2006.
- [16] Xie, Y., Lord, D. and Zhang, Y., Predicting motor vehicle collisions using Bayesian Neural Network Models: an empirical analysis. *Accident Analysis & Prevention*, 39(5), pp. 922-933, 2007. DOI: 10.1016/j.aap.2006.12.014
- [17] Chang, L., Analysis of freeway accident frequencies: negative binomial regression versus Artificial Neural Network. *Safety Science*, 43(8), pp. 541-557, 2005. DOI: 10.1016/j.ssci.2005.04.004
- [18] Li, X., Lord, D., Zhang, Y. and Xie, Y., Predicting motor vehicle crashes using Support Vector Machine Models. *Accident Analysis & Prevention*, 40(4), pp. 1611-1618, 2008. DOI: 10.1016/j.aap.2008.04.010
- [19] Abdel-Aty, M. and Abdelwahab, H., Predicting injury severity levels in traffic crashes: a modeling comparison. *Journal of Transportation Engineering*, 130(2), pp. 204-210, 2004. DOI: 10.1061/(ASCE)0733-947X(2004)130:2(204)
- [20] Abdelwahab, H. and Abdel-Aty, M., Artificial neural networks and logit models for traffic safety analysis of toll plazas. *Transportation Research Record*, 1784, pp. 115-125, 2002. DOI: 10.3141/1784-15
- [21] Genders, W. and Razavi, S., Using a deep reinforcement learning agent for traffic signal control. *arXiv*, 2016.
- [22] Genders, W. and Razavi, S., Evaluating reinforcement learning state representations for adaptive traffic signal control. *Procedia Computer Science*, 130, pp. 26-33, 2018. DOI: 10.1016/j.procs.2018.04.008
- [23] Karlaftis, M. and Vlahogianni, E., Statistical methods versus neural networks in transportation research: differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), pp. 387-399, 2011. DOI: 10.1016/j.trc.2010.10.004
- [24] Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R. and Dera, D., Machine learning in transportation data analytics. In: Chowdhury, M., Apon, A. and Dey K., Eds. *Data analytics for intelligent transportation system*, Elsevier, 2017, pp. 283-307. DOI: 10.1016/B978-0-12-809715-1.00012-2
- [25] Ross, T., The synthesis of intelligence - its implications. *Psychological Review*, 45(2), pp. 185-189, 1938. DOI: 10.1037/h0059815
- [26] Samuel, A., Some studies in Machine Learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), pp. 210-229, 1959. DOI: 10.1147/rd.33.0210
- [27] Abduljabbar, R., Dia, H., Liyanage, S. and Bagloee, S., Applications of artificial intelligence in transport: an overview. *Sustainability*, 11(1), pp. 189-190, 2019. DOI: 10.3390/su11010189
- [28] Khan, A., Baharudin, B., Lee, H. and Khan, K., A review of Machine Learning algorithms for text-documents classification. *Journal of Advances in Information Technology*, 1(1), pp. 4-20, 2010. DOI: 10.4304/jait
- [29] Agrawal, R., Imielinski, T. and Swami, A., Mining association rules between sets of items in large databases. *Proceedings of ACM SIGMOD Conference*, Washington, D.C., 1993, pp. 207-216

- [30] Karlik, B., Machine learning algorithms for characterization of EMG signals. *International Journal of Information and Electronics Engineering*, 4(3), pp. 189-194, 2014. DOI: 10.7763/ijee.2014.v4.433
- [31] Pineda-Jaramillo, J.D., Insa, R. and Martínez, P., Modeling the energy consumption of trains applying neural networks. *Journal of Rail and Rapid Transit*, 232(3), pp. 816-823, 2017. DOI: 10.1177/0954409717694522
- [32] Bishop, C., *Neural networks for pattern recognition*, Oxford, Clarendon Press, 1995.
- [33] McCulloch, W. and Pitts, W., A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), pp. 115-133, 1943. DOI: 10.1007/BF02478259
- [34] Lamounier, E., Soares, A., Andrade A. and Carrijo, R., A virtual prosthesis control based on neural networks for EMG pattern classification. *Proceedings Artificial Intelligence and Soft Computing*, 2002.
- [35] Soares, A., Adriano, A., Lamounier, E. and Carrijo, R., The development of a virtual myoelectric prosthesis controlled by an EMG pattern recognition system based on neural networks. *Journal of Intelligent Information Systems*, 21(2), pp. 127-141, 2003. DOI: 10.1023/A:1024758415877
- [36] Karlik, B., A Fuzzy clustering Neural Network architecture for multi-function upper limb prosthesis. *IEEE Transactions on Biomedical Engineering*, 50(11), pp. 1255-1261, 2003. DOI: 10.1109/tbme.2003.818469
- [37] Liu, Z. and Luo, Z., Hand motion pattern classifier based on EMG using wavelet packet transform and LVQ neural networks. *IEEE International Symposium on IT in Medicine and Education*, Xiamen, 2008. DOI: 10.1109/itme.2008.4743817
- [38] Cantarella, G. and de Luca, S., Modeling transportation mode choice through artificial neural networks. *Fourth International Symposium on Uncertainty Modeling and Analysis (ISUMA)*, College Park, US, 2003. DOI: 10.1109/isuma.2003.1236145
- [39] Celikoglu, H., Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. *Mathematical and Computer Modelling*, 44(7), pp. 640-658, 2006. DOI: 10.1016/j.mcm.2006.02.002
- [40] Zhao, D., Shao, C., Li, J., Dong, C. and Liu, Y., Travel mode choice modeling based on improved probabilistic neural network. *Seventh International Conference on Traffic and Transportation Studies*, Kunming, China, 2010. DOI: 10.1061/41123(383)65
- [41] Omrani, H., Charif, O., Gerber, P., Awasthi, A. and Trigano, P., Prediction of individual travel mode with evidential Neural Network model. in *Transportation Research Record*, 2399(1), pp. 1-8, 2013. DOI: 10.3141/2399-01
- [42] Lai, X. and Schonfeld, P., Optimizing rail transit alignment connecting several major stations. *Transportation Research Board 89th Annual Meeting*, Washington, D.C., 2010.
- [43] Jha, M., Schonfeld, P. and Samanta, S., Optimizing rail transit routes with genetic algorithms and geographic information systems. *Journal of Urban Planning and Development*, 133(3), pp. 161-171, 2007. DOI: 10.1061/(ASCE)0733-9488(2007)133:3(161)

- [44] Pineda-Jaramillo, J.D., Modelo de optimización del consumo energético en trenes mediante el diseño geométrico vertical sinusoidal y su impacto en el coste de la construcción de la infraestructura. Tesis PhD, Departamento de Ingeniería e Infraestructura del Transporte, Universitat Politècnica de València, España, 2017. DOI: 10.4995/Thesis/10251/90546
- [45] Samanta, S. and Jha, M., Modeling a rail transit alignment considering different objectives. *Transportation Research: Part A*, 45(1), pp. 31-45, 2011. DOI: 10.1016/j.tra.2010.09.001
- [46] Pastori, L., Kautbruegger, R. and Budich, J., Generalized transfer matrix states from artificial neural Networks. *Physical Review B*, 99(16), pp. 165123-165134, 2019. DOI: 10.1103/PhysRevB.99.165123
- [47] Banchi, L., Grant, E., Rocchetto, A. and Severini, S., Modelling non-Markovian quantum processes with recurrent neural Networks. *New Journal of Physics*, 20, pp. 123030-123042, 2018. DOI: 10.1088/1367-2630/aaf749
- [48] Iten, R., Metger, T., Wilming, H., del Río, T. and Renner, R., Discovering physical concepts with Neural Networks. *eprint arXiv:1807.10300*, 2018.
- [49] Xin, T., Lu, S., Cao, N., Anikeeva, G., Lu, D., Li, J., Long, G. and Zeng, B., Local-measurement-based quantum state tomography via Neural Networks. *eprint arXiv:1807.07445*, 2018.
- [50] Weinstein, S., Neural Networks as "hidden" variable models for quantum systems. *eprint arXiv:1807.03910*, 2018.
- [51] Stergiou, C. and Siganos, D., Neural Networks. Department of Computing - Imperial College London, 1996.
- [52] Fritsch, J., Modular neural networks for speech recognition. PhD Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, USA, 1996.
- [53] Sacco, D., Motta, G., You, L., Bertolazzo, N., Carini, F. and Ma, T., Smart cities, urban sensing, and big data: mining geo-location in social networks, in: Liu, X., Anand, R. and Xiong, G., (Eds), *Big data and smart service systems*, Zhejiang University Press, 2017, pp. 59-84. DOI: 10.1016/b978-0-12-812013-2.00005-8
- [54] Liang, X. and Wang, G., A convolutional Neural Network for transportation mode detection based on smartphone platform. *IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, Orlando, 2017. DOI: 10.1109/mass.2017.81
- [55] Sak, H., Senior, A. and Beaufays, F., Long short-term memory recurrent Neural Network architectures for large scale acoustic modeling. *Conference of the International Speech Communication Association (INTERSPEECH)*, Singapore, 2014.
- [56] Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning*, MIT press, Boston, USA, 2016.
- [57] Ma, X., Dai, Z., He, Z., Na, J., Wang, Y. and Wang, Y., Learning traffic as images: a deep convolutional Neural Network for large-scale transportation Network speed prediction. *Sensors*, 17(4), pp. 1-16, 2017. DOI: 10.3390/s17040818
- [58] de Oña, J., de Oña, R. and López, G., Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation. *Transportation*, 43(5), pp. 725-747, 2016. DOI: 10.1007/s11116-015-9615-0

- [59] Quinlan, J., C4.5: programs for machine learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1992.
- [60] Breiman, L., Bagging predictors. Machine Learning, 24(2), pp. 123-140, 1996. DOI: 10.1023/A:1018054314350
- [61] Lantz, B., Machine learning with R, Birmingham: Packt Publishing, 2015.
- [62] Hagenauer, J. and Helbich, M., A comparative study of Machine Learning classifiers for modeling travel mode choice. Expert Systems with Applications, 78 pp. 273-282, 2017. DOI: 10.1016/j.eswa.2017.01.057
- [63] Breiman, L., Random forests. Machine Learning , 45(1), pp. 5-32, 2001. DOI: 10.1023/A:1010933404324
- [64] Vapnik, V., The nature of statistical learning theory, Second Ed., Springer Science & Business Media, New York, USA, 2000. DOI: 10.1007/978-1-4757-3264-1
- [65] Cortes, C. and Vapnik, V., Support-Vector networks. Machine Learning , 20(3), pp. 273-297, 1995. DOI: 10.1023/A:1022627411411
- [66] Ben-Hur, A., Horn, D., Siegelmann, H. and Vapnik, V., Support vector clustering. Journal of Machine Learning Research , 2(12), pp. 125-137, 2001. DOI: 10.1162/15324430260185565
- [67] Hair Jr, J., Black, W., Babin, B. and Anderson, R., Multivariate data analysis, Seventh Ed., Pearson, Harlow, UK, 2014.
- [68] Fraley, C. and Raftery, A., How many clusters?. Which clustering method?. Answers via model-based cluster analysis. The Computer Journal, 41(8), pp. 578-588, 1998. DOI: 10.1093/comjnl/41.8.578
- [69] Magidson, J. and Vermunt, J., Latent class models for clustering: a comparison with K-means. Canadian Journal of Marketing Research, 20, pp. 37-44, 2002.
- [70] Karlaftis, M. and Tarko, A., Heterogeneity considerations in accident modeling. Accident Analysis & Prevention , 30(4), pp. 425-433, 1998. DOI: 10.1016/S0001-4575(97)00122-X
- [71] Outwater, M., Castleberry, S., Shiftan, Y., Ben-Akiva, M., Zhou, Y. and Kuppam, A., Attitudinal market segmentation approach to mode choice and ridership forecasting. Structural equation modeling. Transportation Research Record, 1854(1), pp. 32-42, 2003. DOI: 10.3141/1854-04
- [72] Ma, J. and Kockelman, K., Crash frequency and severity modeling using clustered data from Washington state, in: IEEE Intelligent Transportation Systems Conference, Toronto, Canada, 2006. DOI: 10.1109/itsc.2006.1707456
- [73] Depaire, B., Wets, G. and Vanhoof, K., Traffic accident segmentation by means of latent class clustering. Accident Analysis & Prevention , 40(4), pp. 1257-1266, 2008. DOI: 10.1016/j.aap.2008.01.007
- [74] de Oña, J., López, G., Mujalli, R. and Calvo, F., Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. Accident Analysis & Prevention , 51, pp. 1-10, 2013. DOI: 10.1016/j.aap.2012.10.016
- [75] de Oña, R. and de Oña, J., Analyzing transit service quality evolution using decision trees and gender segmentation. WIT transactions on the built environment, 130, pp. 611-621, 2013. DOI: 10.2495/ut130491

- [76] Sprumont, F. and Viti, F., The effect of workplace relocation on individual's activity travel behavior. *Journal of Transport and Land Use*, 11(1), pp. 985-1002, 2018. DOI: 10.5198/jtlu.2018.1123
- [77] Cantelmo, G., Viti, F., Cipriani, E. and Nigro, M., A utility-based dynamic demand estimation model that explicitly accounts for activity scheduling and duration. *Transportation Research Procedia*, 23, pp. 440-459, 2017. DOI: 10.1016/j.trpro.2017.05.025
- [78] Sprumont, F., Astegiano, P. and Viti, F., On the consistency between commuting satisfaction and traveling utility: the case of the University of Luxembourg. *European Journal of Transport and Infrastructure Research*, 17(2), pp. 248-262, 2017. DOI: 10.18757/ejtir.2017.17.2.3193
- [79] Muñoz, C., Córdoba, J. and Sarmiento, I., Airport choice model in multiple airport regions. *Journal of Airline and Airport Management*, 7(1), pp. 1-12, 2017. DOI: 10.3926/jairm.62
- [80] Zhao, X., Yan, X., Yu, A. and Van Hentenryck, P., Modeling Stated preference for mobility-on-demand transit: a comparison of Machine Learning and logit models. *arXiv :Doi: 1811.01315*, 2018.
- [81] Shmueli, D., Salomon, I. and Shefer, D., Neural Network analysis of travel behavior: evaluating tools for prediction. *Transportation Research Part C: Emerging Technologies*, 4(3), pp. 151-166, 1996. DOI: 10.1016/S0968-090X(96)00007-1
- [82] Sayed, T. and Razavi, A., Comparison of neural and conventional approaches to mode choice analysis *Journal of Computing in Civil Engineering*, 14(1), pp. 23-30, 2000. DOI: 10.1061/(ASCE)0887-3801(2000)14:1(23)
- [83] Mohammadian, A. and Miller, E., Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. *Transportation Research Record*, 1807(1), pp. 92-100, 2002. DOI: 10.3141/1807-12
- [84] Vythoulkas, P. and Koutsopoulos, H., Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks. *Transportation Research Part C: Emerging Technologies*, 11(1), pp. 51-73, 2003. DOI: 10.1016/S0968-090X(02)00021-9
- [85] Hensher, D. and Ton, T., A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E: Logistics and Transportation Review*, 36(3), pp. 155-172, 2000. DOI: 10.1016/S1366-5545(99)00030-7
- [86] Xie, C., Lu, J. and Parkany, E., Work travel mode choice modeling with data mining: decision trees and neural networks. *Transportation Research Record*, 1854(1), pp. 50-61, 2003. DOI: 10.3141/1854-06
- [87] Andrade, K., Uchida, K. and Kagaya, S., Development of transport mode choice model by using adaptive neuro-fuzzy inference system. *Transportation Research Record*, 1977(1), pp. 295-304, 2006. DOI: 10.1177/0361198106197700102
- [88] Zhang, Y. and Xie, Y., Travel mode choice modeling with Support Vector Machines. *Transportation Research Record*, 2076(1), pp. 141-150, 2008. DOI: 10.3141/2076-16
- [89] Pulugurta, S., Arun, A. and Errampalli, M., Use of artificial intelligence for mode choice analysis and comparison with traditional multinomial logit

- model. *Procedia - Social and Behavioral Sciences*, 104, pp. 583-592, 2013. DOI: 10.1016/j.sbspro.2013.11.152
- [90] Teng, H. and Qi, Y., Detection-delay-based freeway incident detection algorithms. *Transportation Research Part C: Emerging Technologies*, 11(3-4), pp. 265-287, 2003. DOI: 10.1016/S0968-090X(03)00022-6
- [91] Teng, H. and Qi, Y., Application of wavelet technique to freeway incident detection. *Transportation Research Part C: Emerging Technologies*, 11(3-4), pp. 289-308, 2003. DOI: 10.1016/S0968-090X(03)00021-4
- [92] Rasouli, S. and Timmermans, H., Using ensembles of decision trees to predict transport mode choice decisions: effects on predictive success and uncertainty estimates. *European Journal of Transport and Infrastructure Research*, 14(4), pp. 412-424, 2014.
- [93] Tang, L., Xiong, C. and Zhang, L., Decision tree method for modeling travel mode switching in a dynamic behavioral process. *Transportation Planning and Technology*, 38(3), pp. 833-850, 2015. DOI: 10.1080/03081060.2015.1079385
- [94] Zhan, G., Yan, X., Zhu, S. and Wang, Y., Using hierarchical tree-based regression model to examine university student travel frequency and mode choice patterns in China. *Transport Policy*, 45, pp. 55-65, 2016. DOI: 10.1016/j.tranpol.2015.09.006
- [95] Ravi-Sekhar, C., Minal and Madhu, E., Mode choice analysis using random Forrest decision trees. *Transportation Research Procedia*, 17, pp. 644-652, 2016. DOI: 10.1016/j.trpro.2016.11.119
- [96] Cheng, L., Chen, X., de Vos, J., Lai, X. and Witlox, F., Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society*, 14, pp. 1-10, 2019. DOI: 10.1016/j.tbs.2018.09.002
- [97] Omrani, H., Predicting travel mode of individuals by Machine Learning. *Transportation Research Procedia*, 10, pp. 840-849, 2015. DOI: 10.1016/j.trpro.2015.09.037
- [98] Xian-Yu, J., Travel mode choice analysis using Support Vector Machines, in 11th International Conference of Chinese Transportation Professionals (ICCTP), Nanjing, China, 2011. DOI: 10.1061/41186(421)37
- [99] Ding, L. and Zhang, N., A travel mode choice model using individual grouping based on cluster analysis. *Procedia Engineering*, 137, pp. 786-795, 2016. DOI: 10.1016/j.proeng.2016.01.317
- [100] Li, J., Weng, J., Shao, C. and Guo, H., Cluster-Based logistic regression model for holiday travel mode choice. *Procedia Engineering*, 137, pp. 729-737, 2016. DOI: 10.1016/j.proeng.2016.01.310
- [101] Pirra, M. and Diana, M., Classification of tours in the U.S. National household travel survey through Clustering Techniques. *Journal of Transportation Engineering*, 142(6), pp. 1-13, 2016. DOI: 10.1061/(ASCE)TE.1943-5436.0000845
- [102] Molin, E., Mokhtarian, P. and Kroesen, M., Multimodal travel groups and attitudes: a latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice*, 83, pp. 14-29, 2016. DOI: 10.1016/j.tra.2015.11.001
- [103] Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D., Do we need hundreds of classifiers to solve real world classification problems?. *Journal of Machine Learning Research*, 15(1), pp. 3133-3181, 2014.

- [104] Dia, H. and Panwai, S., Evaluation of discrete choice and neural network approaches for modelling driver compliance with traffic information. *Transportmetrica*, 6(4), pp. 249-270, 2010. DOI: 10.1080/18128600903200596
- [105] Dia, H. and Panwai, S., Modelling drivers' compliance and route choice behaviour in response to travel information. *Nonlinear Dynamics*, 49(4), pp. 493-509, 2007. DOI: 10.1007/s11071-006-9111-3
- [106] Nijkamp, P., Reggiani, A. and Tritapepe, T., Modelling inter-urban transport flows in Italy: a comparison between Neural Network analysis and logit analysis. *Transportation Research Part C: Emerging Technologies*, 4(6), pp. 323-338, 1996. DOI: 10.1016/S0968-090X(96)00017-4
- [107] Pineda-Jaramillo, J.D., Black-box model using ANN to reduce energy consumption in railway lines and their impact on infrastructure construction costs. 20th Pan-American Conference of traffic, transportation and logistics engineering (PANAM), Medellín, Colombia, 2018. ISSN: 2711-032X

Notes

J.D. Pineda-Jaramillo, received the BSc. Eng. in Civil Engineering in 2011, the MSc. in Transportation Systems in 2013, both from the Universidad Nacional de Colombia, and the PhD. in Civil Engineering in 2017 from the Technical University of Valencia, Spain, and he is an enthusiast in Artificial Intelligence. Throughout his experience as professional, researcher and lecturer, he has built solid analytic skills in transportation planning, mobility, railway engineering, modeling, and he has developed a strong interest on Data Sciences for transportation research. He is currently working on data-driven and model-based approaches for transportation research. ORCID: 0000-0002-4657-7521

How to cite: Pineda-Jaramillo, J.D, A review of Machine Learning (ML) algorithms used for modeling travel mode choice. *DYNA*, 86(211), pp. 32-41, October - December, 2019.