A Proposal for Carpooling based on User's Preferences

César Mendes Santos, Salvador Ibarra M., José A. Castán R., Jesús David Terán V., Julio Laria M., Mayra G. Treviño B., Julissa Pérez C. and Emilio Castán R.

Abstract— **Traffic congestion is a serious problem in many urban areas around the world, which generates endless negative impacts on society and the economy, among which are: air pollution due to carbon dioxide emissions, waste of fuel and increase travel time. In this sense, the Carpooling initiative, which is the dynamic in which a driver shares his car with one or more additional passengers who have a similar destination, emerges as one of the most effective solutions to deal with the problems generated by the vehicular congestion. This paper shows the development of an intelligent system for the use of Carpooling which allows users to have a tool to carry out this practice, which, unlike existing ones, allows access at any time and in any place to the user's data such as their current location, their destination, their tastes and personal preferences. To select the most optimal route, the ideal driver and the companions, the system uses one of the most used models in supervised learning and in data mining applications, which is the ID3 decision tree technique and as a form of backup in the decision making the system makes use of the Naive Bayes technique. The obtained results, after the application of the developed system, showed an effectiveness of 86.86% which demonstrate that when using the automatic learning techniques in Carpooling, results can be obtained in a similar way to how the user would do it.**

Index Terms—**Carpooling, user's preference, ideal selection of passenger.**

I. Introduction

DUE to the constant growth or urban population cities face enormous challenges related to transportations, traffic congestion, control of carbon emissions pollution of the environment and prevention of car accidents. As an answer to these problems, the Carpooling initiative, which is the dynamic in which a driver shares his car with one or more additional passengers who have a similar destination is an environmentally friendly transportation strategy and is one of the most effective solutions to reduce traffic congestion [1].

Manuscript received March 03, 2017; revised March 23, 2017.

S. Ibarra, J. A. Castán, J. Laria, J. Terán, M. G. Treviño, J. Perez are members of the Computer Technology Group at the Engineering School at the Autonomous University of Tamaulipas, Mexico. (sibarram@uat.edu.mx).

C. Mendes is a PhD Student at the Autonomous University of Tamaulipas, Mexico.

E. Castán is professor at the Electric and Electronic Department at the TecNM/Madero Institute of Technology.

Carpooling strategy was proposed in the 1940s by the United States government during the Second World War, when the shortage of fuel and rubber for the manufacture of tires, forced the inhabitants to make a more reasonable use of private vehicles for personal transportation [2]. Currently it is one of the most effective practices to reduce vehicular traffic and fuel consumption, likewise Carpooling, is used as a means of transport among the inhabitants of a modern city.

According to the National Report on Urban Mobility in Mexico 2014-2015 [3], the average occupancy rate of trips in personal vehicles is 1.2 persons per kilometer traveled. If a regular vehicle can transport a maximum of 5 people and only transports 1.2 people, then 76% of the transport capacity is wasted during the trip. Therefore, the implementation of the Carpooling service, the car occupancy rate could be substantially increased by reducing the number of empty seats in these vehicles. If consider that only in Mexico, where 25 million personal vehicles circulate daily, the implementation of the Carpooling service would reduce this number of vehicles, having a positive effect on the use of roads and reducing emissions of pollutants to the environment.

To provide a Carpooling service, various systems have been proposed based on both desktop, web and mobile applications [4], [5], [6], [7], [8], [9]. These services are based on an application in which the user's information are transmits to an online public platform in which an interface is provided between the passengers that are looking for a shared trip and the drivers that offer their vehicle to make the trip. Service users can search all published travel plans and should contact the service provider they are interested in. Some of these systems provide basic functions to carry out the Carpooling, with the possibility of sending a service request in a certain place, date and time, and then find suitable travel companions to satisfy the demand.

In [10] and [11] is made a proposal for a Carpooling system based on an application for mobile devices which allows users to have an easy and fast access to the system no matter where they are. This application allows the users to perform a pairing between drivers and passengers who have a similar route. The process of selecting fellow travelers is based on the origin and destination place of the users requesting the service. These data are processed and evaluated by a system which use of a genetic algorithm to give the best solution based on the location of each requesting user.

In [12] a recommendation system for traveling companions based on data mining is proposed. To perform this procedure, users must allow the application to collect information from their travel routes in such a way that the mobile device, using GPS, monitors each user, said information is stored in a database that will serve the system for decision making. The system processes each of the stored routes and compares which are the routes with the highest degree of similarity and based on this recommends the possible travel companions between passengers and drivers.

In [13] a system for the city of Shenzhen in China is proposed, which allows taxi drivers to implement Carpooling for the large number of passengers going and come to the airport. For these, a mobile application was generated that is used by the passengers and an embedded system is place in the taxi. In this sense the passengers request a taxi through the mobile application, these requests are processing by a server which generates the optimal route for the taxi so that it can pass through the largest number of passengers who are closer, allowing sharing the vehicle and thereby reduce the amount of travel and pollution issued by each vehicle.

In [14] are based on sharing a vehicle but emphasizing the privacy of users. In this case when the drivers and passengers make the service request, the data is sent to the server and the system evaluates the place of origin and destination of each user, and a possible travel route for the driver is generated, then the passengers are selected taking in account the closer distance from the travel route, and once the passengers are selected, pseudo-randomly, the meeting points between the driver and the passengers are generated, thus maintaining the privacy of users regarding their meeting place.

All the works cited above, provide some solution for the selection of travel companion, in the practice of Carpooling, however, these systems are inefficient for the user that requires a pairing service in real time and that suits their needs tastes and preferences. Therefore, in the present work we propose an Intelligent Carpooling System (SINCAR) with a cloud-based architecture which is accessed through a mobile application and in which the personal tastes and preferences of the driver and the users are considered.

II. Methodology

The Intelligent Carpooling System developed, is based on the use of Cloud Computing. As described in [15] the term cloud computing refers to the abstraction of computers, resources and web-based services that system developers can use to implement complex web systems. Often, these cloud-based resources are considered virtual, which means that if a system or solution needs more resources, such as processors or disk space, the resources can simply be added to the demand and, transparently, to the applications that use them.

On the other hand, SINCAR was developed in such a way that users and drivers send their Carpooling service requests with their current location, as well as their personal preferences, which will be processed by SINCAR who making decisions for the pairing between the driver and the passengers, as well as to define the optimal route. To select the most optimal route, an ID3 algorithm with the attributes generated by the users is applied, allowing SINCAR to adapt to the tastes of each driver and each user. As the decision model learns, the system arrive a point where no longer depend by the user to determine a response and it can make decisions in a way very similar to how the user of the system would.

Decision trees are one of the most used models in supervised learning and in data mining applications [16]. A decision tree is a set of conditions organized in a hierarchical structure, in such a way that the final decision to be made can be determined by following the conditions that are fulfilled from the root of the tree to some of its leaves. They are classification models easy to understand. Its application domain is not restricted to a specific field but can be used it in different areas [17], from medical diagnostic applications to weather forecast systems or games such as chess.

One of the algorithms for generating the decision trees is the ID3, developed by J. Ross Quinlan [18] during the decade of the 70's. This algorithm employs a top-down search in among the possible attributes of the decision tree, to find the best one that classifies. The choice of the best attribute is made according to the information gain that is based on the entropy concept of information theory and which is calculated using formula (1) [19].

$$H(x) = \sum P(x_i) \operatorname{Log} P(x_i)^{-1} \qquad (1)$$

Where:
P (xi) - is the probability that the event xi occurs
Log P (xi) – is size to represent the occurrence of xi

A learning algorithm like ID3 needs a database of examples known as "training examples". One of the algorithm characteristics is to generalize from a series of patterns that are provided by the user which are included in the database. This algorithm can generalize and learn patterns that were not provided in database. Therefore, to train the ID3 algorithm it is necessary to have a training database with enough examples that can be used to classify previously unknown patterns.

To determine the number of cases with which the decision tree should be trained, the formula (2), from statistics, was used to calculate the sample size for finite populations.

$$n = \frac{N\,Z^2 pq}{(N-1)E^2 + Z^2 pq} \qquad (2)$$

Where:
n: Sample size.
N: Universe.
p: Probability in favor.
q: Probability against.
E: Estimation error.
Z: Confidence level.

The ID3 algorithm is the central part of the proposed Intelligent Carpooling System (SINCAR), however, if there is not enough examples in database to provide a solution using this algorithm, the model will provide a solution based on the implementation of the Naive Bayes technique. Therefore, the implementation of the Naive Bayes technique in the SINCAR is proposed as a rectification in the decision making of the ID3 algorithm, when it is haven't enough examples in database to provide a solution using this algorithm.
In this sense, the Naive Bayes technique has the advantage of being a probabilistic classifier that can be help in the resolution of this type of inconvenience. Its operation is based on the use of frequency to calculate the success probability in new cases. This technique use formula (3). For the Naive Bayes training process, the same cases that were used by the ID3 decision tree were used.

$$P(C|X) = \frac{P(C|X)P(C)}{P(X)} \qquad (3)$$

Where:
C: Previous data set.
X: New data to be evaluated.
P (C | X): Degree of belonging.
P (C): Total probability of the class.
P (X): Evidence.

III. The Approach

The general structure of the Intelligent Carpooling System (SINCAR) developed is shown in Figure 1. This system is divided into two modules: the Customer Module (MC) and the Cloud Computing Carpooling Module (M3C). The connections between the MC and the M3C can be carried out by any mobile device with internet connection through Wi-Fi, CDMA, UMTS, 3G, 4G or LTE [16], [17].

The MC is the interface through which the users interact with the system. This module transmits the user's personal information, ie the user's location, destination and personal preferences and travel requests to the M3C. On the other hand, the MC also shows the information sent from the M3C, ie the data of the assigned users as well as the generated routes. The MC developed to be compatible with the IOS and Android operating systems, since both have the GPS module and internet access needed to the correct operation of the system.

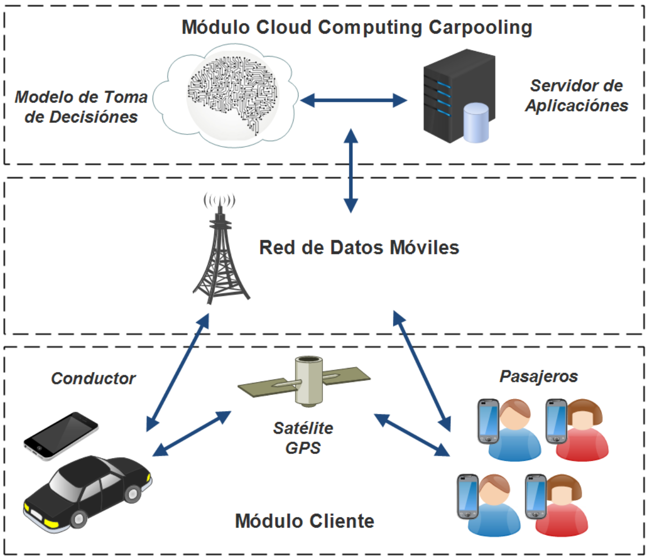

Figure 1. Structure of the Intelligent Carpooling System (SINCAR).

The MC is the interface through which the users interact with the system. This module transmits the user's personal information, ie the user's location, destination and personal preferences and travel requests to the M3C. On the other hand, the MC also shows the information sent from the M3C, ie the data of the assigned users as well as the generated routes. The MC developed to be compatible with the IOS and Android operating systems, since both have the GPS module and internet access needed to the correct operation of the system.

In the M3C module, the drivers and passengers are assignment based on the travel requests of drivers and passengers, their locations, and their personal preferences. These results are send to the MC module which continue interact with the assigned users until they arrive to destination.

The operation of the M3C is controlled by the SINCAR algorithm whose general structure is shown in Figure 2. The SINCAR algorithm is divided into three stages, the first stage is the reception of carpooling requests from the users, then in the second stage these requests are processed by the Decision-Making Model (DMM), and end stage, the best travel route for the requests is generates. The DMM was developed by three stages. In the first stage the user's information's are processed by classifying them by origin and destination, such data is filtering and sending to the next stage only the requests that have a similar origin-destination, thus avoiding spending computational resources on requests that do not present data in common. Once the requests have been classified, they pass to the second stage, which is the selection of users. This stage is the central part of the DMM, since in this stage the preferences of the users are evaluated

ISBN: 978-988-14048-6-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

to define the best result that meets the criteria that are consistent with that have stored in data base.

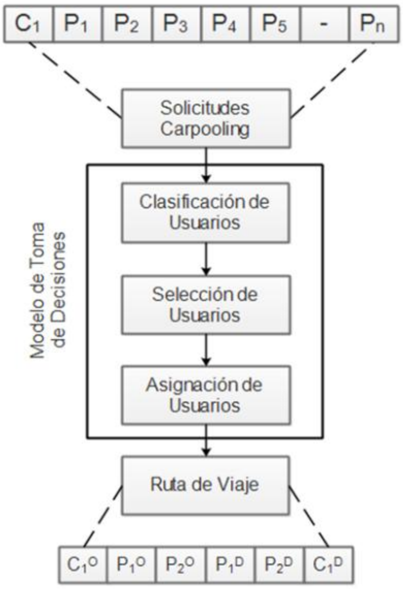

Figure 2. General structure of the SINCAR algorithm.

The ID3 decision tree technique is used for the assignment of users, since it is a supervised learning algorithm that is commonly used to train different types of applications where a specific amount of possible variables are input and as the result a dichotomous response is obtained that allow the model to decide which users are compatible to carry out a shared trip, based on their personal preferences.

The attributes used in this technique are shown in table 1. Each one is associated to a set of labels and in some cases to a set of a range of values. Such information allows differentiating between the possible states for each attribute. The choice of the best attribute is made according to the gain of information that is based on the concept of entropy which is calculated using formula (1).

Table 1. Attributes used in the decision tree technique.

Attributes	
Age	18 - 24 25 - 40 41 - 50 50 - 64
Smoker	Si No
Sex	Male Female
Distance	Near Moderate Far
Music	Little Regular A lot

The implementation of the ID3 algorithm allows to obtain a solution based on the previous experiences of each user, adapting to the tastes and preferences of these, trying to make the decisions in the same way as the user would. However, if not there is enough information to provide a solution using this algorithm ID3, the model will provide a solution based on the Naive Bayes technique.

To determine the number of cases to train the decision tree the formula (2) was used with the 50% of probability in favor, 5% of the margin error and the 95% of confidence level, obtaining a result of 105 cases. This is the number of cases which the ID3 need to obtain a result like the reality. Each user will have its unique decision tree, because the tastes and personal preferences are different. Figure 3 shows the tree generated with the training cases made for a specific driver.

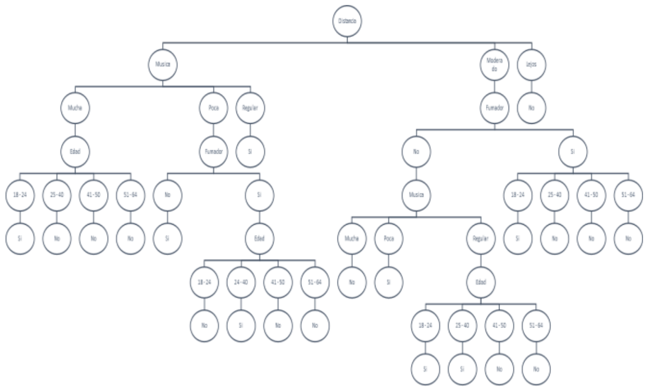

Figure 3. Tree obtained for a specific driver.

One of the drawbacks of ID3 is that, if it does not have enough number of training cases, as is the case of the new user, it will not know what decision to take and thus hindering the decision-making process. Therefore, as a form of backup or rectification in the decision making of the ID3 algorithm, we use the Naive Bayes technique defined according to the formula (3). For the Naive Bayes training process, were used the same cases that in the ID3 decision tree.

The third stage of the DMM is the user assignment (drivers and passengers), for this purpose only the users with the more amount of tastes and preferences compatible are assigned. Finally, the SINCAR algorithm generates the best travel route considering the origin and destination of the users that were selected.

To evaluate the performance of SINCAR, in a scenario of real tests, with different types of vehicles and user, a case study was made at the Autonomous University of Tamaulipas, South University Center, with three different test scenarios as shown in Table 2.

For each scenario 1000 experiments were made, where the driver proposed to make the trip and the SYNCAR generates the best travel route considering the origin and destination of the users and their tastes and preferences. In the all scenarios the destination was the same, the South

University Center but towards three different accesses, as shown in Figure 4.

Table 2. Test scenarios.

Escenario 1	
Number of seats available in the vehicle	4 seats
Origen	Calle 4. Col. Jardín 20 de Noviembre. Madero
Destination	Centro Universitario Sur Acceso 1
Distance between origin and destination	8.4 Km
Average travel time without Carpooling	12 min
Driver	Student
Escenario 2	
Number of seats available in the vehicle	3 seats
Origen	Calle Alfredo Gomez Vega. Col. Asunción Avalos. Madero
Destination	Centro Universitario Sur Acceso 2
Distance between origin and destination	9.6 Km
Average travel time without Carpooling	10 min
Driver	Teacher
Escenario 3	
Number of seats available in the vehicle	3 seats
Origen	Calle Huasteca. Col. Lomas de Infonavit. Tampico
Destination	Centro Universitario Sur Acceso 3
Distance between origin and destination	6.2 Km
Average travel time without Carpooling	8 min
Driver	Teacher

Figure 4. Aerial view of access to the University Campus.

For each 1000 experiments carried out in each scenario, the level of effectiveness of the decisions taken by SINCAR was measured. For this, each driver was asked if he agreed with the assigned users (% user) and each user if they agreed with the assigned driver (% driver). Finally, both results are weighted (% trip), to obtain a final level of effectiveness of the decisions taken by SINCAR. Tables 3, 4 and 5 show the results of 10 experiments taken at random among the 1000 experiments performed for each scenario respectively.

For the scenario 1 (see table 3) in experiment 1, the driver did not agree with 1 of the users (75% users) proposed by the system and two of the four users (50% driver) were not agree to the assigned driver, therefore, a total of 62.5% of effectiveness was obtained for this experiment. As a result of the 10 experiments carried out in scenario 1, the level of agree by the driver of the users assigned by SINCAR was 85.84% and the driver agree by the users was 86.67%, resulting in a system effectiveness of 85.01%.

Table 3. Results of the evaluation of the effectiveness of SINCAR for scenario 1.

No.	Seats in the car	Asigned seats	% driver	% user	% trip
1	4	4	50.0	75.0	62.5
2	4	4	75.0	75.0	75.0
3	4	4	100.0	100.0	100.0
4	4	3	66.7	100.0	83.4
5	4	3	100.0	66.7	83.4
6	4	4	50.0	75.0	62.5
7	4	3	66.7	100.0	83.4
8	4	3	66.7	100.0	83.4
9	4	4	100	100.0	100.0
10	4	3	100	66.7	83.4
Average			77.51	85.84	81.7

For the scenario 2 (see table 4) the level of agree by the driver of the users assigned by SINCAR was 80.01% and the driver agree by the users was 88.34%, resulting in a system effectiveness of 84.2%.

Table 4. Results of the evaluation of the effectiveness of SINCAR for scenario 2

No.	Seats in the car	Asigned seats	% driver	% user	% trip
1	3	3	100	100	100
2	3	3	100	66.7	83.4
3	3	3	66.7	100	83.4
4	3	3	66.7	100	83.4
5	3	3	100	66.7	83.4
6	3	1	100	0	50
7	3	2	100	100	100
8	3	2	50	100	75
9	3	3	100	100	100
10	3	3	100	66.7	83.4
Average			88.34	80.01	84.2

For the scenario 3 (see table 5) the level of agree by the driver of the users assigned by SINCAR was 83.34% and

ISBN: 978-988-14048-6-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the driver agree by the users was 86.67%, resulting in a system effectiveness of 85.01%.

Table 5. Results of the evaluation of the effectiveness of SINCAR for scenario 3.

No.	Seats in the car	Asigned seats	% driver	% user	% trip
1	3	3	100	100	100
2	3	3	100	100	100
3	3	2	50	100	75
4	3	2	100	100	100
5	3	2	100	50	75
6	3	1	100	100	100
7	3	3	66.7	66.7	66.7
8	3	3	100	66.7	83.4
9	3	3	100	100	100
10	3	2	50	50	50
Average			86.67	83.34	85.01

The average results of the 1000 experiments performed for the three scenarios showed a level of acceptance by the driver of the users assigned by SINCAR of 85.24% and the acceptance of the driver by the users was 88.47%, resulting in an effectiveness of the 86.86% system.

IV. Conclusions

This paper shows the development of an intelligent system for the use of Carpooling which, unlike existing ones, allows to the user access at any time and in any place any place. In addition, the system developed allows the user to enter not only their current location but also their destination, their tastes and personal preferences. To select the most optimal route, the ideal driver and the companions, the system uses the ID3 decision tree technique which is one of the most used models in supervised learning and in data mining applications. As a form of backup in the decision making of the ID3 algorithm, the system makes use of the Naive Bayes technique.

The results obtained after 1000 experiments, showed an effectiveness of 86.86% in the decisions taken by the developed system. This shows that by using machine learning techniques, can be obtained results in a similar way as the user would. Since the system is more used, it more learns, and this facilitates obtaining better results in the practice of Carpooling.

References

[1] E. Ferguson, "Demographics of Carpooling," *Transp. Res. Rec. 1496*, pp. 142–150, 1990.

[2] E. Ferguson, "The rise and fall of the American carpool: 1970–1990," *Transportation (Amst).*, vol. 24, pp. 349–376, 1997.

[3] ONU-Hábitat, "Reporte Nacional de Movilidad Urbana en México 2014-2015," *Rep. Glob. en Asentam. Humanos*, p. 100, 2015.

[4] E. Nechita, G.-C. C. Crişan, S.-M. M. Obreja, and C.-S. S. Damian, "Intelligent carpooling system: A case study for bacău metropolitan area," *Intell. Syst. Ref. Libr.*, vol. 107, pp. 43–72, 2016.

[5] R. Aissaoui and O. I. Houssaini, "CARPOOLING APPLICATION : KwiGo," no. April, 2015.

[6] A. Dixit, S. Bora, S. Chemate, and N. Kolpekwar, "Real-Time Carpooling System for Android Platform," *Int. J. Eng. adn Innov. Technol.*, vol. 2, no. 6, pp. 436–437, 2012.

[7] K. Chibber, "BlaBlaCar is building a global transportation network out of empty car seats," *Quartz*, 2014. [Online]. Available: http://qz.com/239163/blablacar-is-building-a-global-transportation-network-out-of-empty-car-seats/.

[8] Y. Nalawade, V. Waghmare, and P. Waghmare, "Implementation of Dynamic Carpooling System on Android Platform," *Int. J. Innov. Res. Adv. Eng.*, vol. 2, no. 2, pp. 2014–2016, 2015.

[9] N. Bachhav, P. Malode, N. Bhujbal, and S. Jawale, "Dynamic Ride-Sharing Application on Android Platform," *Int. J. Innov. Adv. Comput. Sci.*, vol. 4, no. March, pp. 53–56, 2015.

[10] Chih-Hsiang Lin, Ming-Kai Jiau, and Shih-Chia Huang, "A cloud computing framework for real-time carpooling services," *6th Int. Conf. New Trends Inf. Sci. Serv. Sci. Data Min.*, pp. 266–271, 2012.

[11] A. Shete, V. Bhandare, L. Londhe, and P. B. Mali, "Intelligent Carpooling System," *Int. J. Comput. Appl.*, vol. 118, no. 4, pp. 975–8887, 2015.

[12] W. He, K. Hwang, and D. Li, "Intelligent carpool routing for urban ridesharing by mining GPS trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2286–2296, 2014.

[13] D. Zhang *et al.*, *Carpooling Service for Large-Scale Taxicab Networks*, vol. 12, no. 3. 2016.

[14] P. Goel, L. Kulik, and K. Ramamohanarao, "Optimal Pick up Point Selection for Effective Ride Sharing," *IEEE Trans. Big Data*, vol. PP, no. 99, p. 1, 2016.

[15] K. Jamsa, *Cloud Computing*. Jones & Bartlett Learning, 2012.

[16] Gehrke, J., Loh, W.-Y. & Ramakrishnan, R. (1999b). Classification and regression: money can grow on trees. [1]Tutorial notes for ACM SIGKDD 1999 international conference on Knowledge Discovery and Data Mining

[17] Piatetsky-Shapiro, G., editor (2001): KDnuggets News, Vol. 1, No. 11, Item2, May 29, 2001

[18] Quinlan, J.R. 1993. Learning Efficient Classification Procedures and Their Application to Chess Games, En R. S. Michalski, J. G. Carbonell, & T. M. Mitchells (Eds.) Machine Learning, The Artificial Intelligence Approach. Morgan Kaufmann, Vol. II, Capítulo 15, páginas 463-482, EE.UU.

[19] Mitchell, T.M. Machine Learning (McGraw-Hill, (1997) Cap. 3: "Decision tree learning"

[20] Y. Huang, Y. Chen, Y. T. Hou, W. Lou, and J. H. Reed, "Recent Advances of LTE / WiFi Coexistence in Unlicensed Spectrum," *IEEE Netw.*, vol. PP, no. 99, pp. 1–7, 2017.

ISBN: 978-988-14048-6-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)