Received: 5 April 2022 | Revised: 7 October 2022 | Accepted: 15 October 2022

DOI: 10.1111/exsy.13177

ORIGINAL ARTICLE

Expert Systems WILEY

# Application of choice models in tourism recommender systems

Ameed Almomani[1] | Paula Saavedra[1] | Pablo Barreiro[1] | Roi Durán[1] | Rosa Crujeiras[2] | María Loureiro[3] | Eduardo Sánchez[1]

[1]CITIUS, University of Santiago de Compostela, Santiago, Spain

[2]School of Mathematics, University of Santiago de Compostela, Santiago, Spain

[3]School of Business, University of Santiago de Compostela, Santiago, Spain

**Correspondence**
Eduardo Sánchez, CITIUS, University of Santiago de Compostela, Santiago, Spain.
Email: eduardo.sanchez.vila@usc.es

**Funding information**
Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia, Grant/Award Number: ED431G/08; EMALCSA, Grant/Award Number: CSC-14-13; Ministerio de Ciencia e Innovación, Grant/Award Number: TIN2014-56633-C3-1-R; Ministerio de Economía, Industria y Competitividad, Gobierno de España, Grant/Award Number: MTM2013-41383P; European Regional Development Fund

## Abstract

Choice models (CM) are proposed in the field of tourism recommender systems (TRS) with the aim of providing algorithms with both a theoretical understanding of tourist's motivations and a certain degree of transparency. The goal of this work is to overcome some of the limitations of current state-of-art algorithms used in TRSs by providing: (1) accurate preferences, which are learnt from user choices rather than from ratings, and (2) interpretable coefficients, which are achieved by means of the set of estimated parameters of CM. The study was carried out with a gastronomic data set generated in an ecological experiment in the tourism domain. The performance of CM has been compared with a set of baseline algorithms (rating-based and ensembles) by using two evaluation metrics: precision and DCG. The CM outperformed the baseline algorithms when the size of the choice set was limited. The findings suggest that CM may provide an optimal trade-off between theoretical soundness, interpretability and performance in the field of TRS.

KEYWORDS

artificial intelligence, choice models, ensembles, knowledge engineering, recommender systems, tourism

## 1 | INTRODUCTION

Tourism is a strategic business domain which manages the movement of people to destinations outside their usual environment and plays a key role in the economic and social development of many countries. In the last decade, the digitalization of the sector has been crucial to reach new markets and offer new experiences to tourists. In this field, intelligent or smart tourist recommender systems (TRS) have become mainstream techniques for suggesting personalized plans, products and information to end users.

The field of Machine Learning has played an important role in providing a number of recommendation techniques that work at the back-end of TRSs (Borràs et al., 2014; Hamid et al., 2021; Kzaz et al., 2018). The most relevant approaches are: content, collaborative, context and ensemble-based models. Content-based algorithms work both with item attributes and past user experiences to learn a profile of preferences for each decision-maker (Burke et al., 2011; Harman, 1995). Tourist profiles are built based on demographic or location-based information which is used to estimate the relevancy of point of interests (Santos et al., 2019). Collaborative-based algorithms, on the other hand, require users to rate items, which are later used to build memory and model-based approaches to predict the rating of any new user-item interaction (Burke et al., 2011; Resnick et al., 1994). Since its arrival, the collaborative approach has been widely adopted as a way of removing the need to manage specific domain information, and ratings have become the key data required to fuel rating-based algorithms. While content and collaborative approaches are based on user-item interaction, context-based techniques include contextual information that informs about the specific circumstances of such interaction (Adomavicius & Tuzhilin, 2011). Objects and points of interest can be used as contextual information to suggest optimal plans for tourist (Le & Pishva, 2016). Nowadays, after the impact of the Netflix prize, ensemble models have become very popular to develop recommender systems. Ensemble learning is a paradigm that aggregates instances of weak algorithms (aka learners) to produce more accurate predictions than those provided by single learners (Friedman et al., 2009; Polikar, 2006). The most popular methods proposed for efficient aggregation of learners (Friedman et al., 2009; Polikar, 2006) are bagging, boosting and random forest. A hybrid ensemble, which combines learners of different nature, has been successfully applied to recommend tourist routes based on location-tagged data (Wan et al., 2018).

The main drawbacks of the techniques applied so far in the development of TRSs are: the lack of a theoretical background to understand the underlying motivational factors conditioning the tourist decision-making, and their interpretability to explain the recommendations. The study and identification of tourist's motivation is a central element of the so called push-pull models in tourism (Crompton, 1979; Dann, 1976; Pestana et al., 2020). They are based on the notion that tourists make choices according to a set of needs and motivations that push them to travel, while tourist items have a set of desirable characteristics that attract them. So, the first pillar to develop a sound TRSs should be to understand and learn those push motivations that may explain both tourist's preferences and behaviours. The machine learning approaches apply different shortcuts in order to solve this problem. All rating-based models learn preferences from ratings considering a strong relationship between them: in memory-based approaches, it is assumed that decision-makers with similar ratings will have similar tastes; and in model-based techniques, ratings are assumed to be the result of a matching between latent factors in an item and the decision-maker's preferences about those factors. The problem is the absence of experimental evidence supporting these assumptions, which makes the accuracy of the learnt tastes/preferences unclear. Furthermore, ensemble-based solutions, while successful in terms of accurate predictions, come at a cost of complexity and an opaque nature that makes the recommendations difficult to explain.

Our work focuses on providing a theoretical background to the algorithms behind tourism recommender systems to alleviate these problems. The contribution to the field of TRSs can be summarized as follows:

- The application of choice models (CM) as a tool to learn tourist's preferences with a sound methodology.
- The exploration of the potential of CM by comparing them with algorithms used in TRSs: (1) advanced rating-based algorithms and (2) ensemble strategies.

The paper is organized in the following way. In the Related Work section, an overview of CM and other choice-based strategies are reviewed. In the Background section, the recommendation problem is described as a choice problem and the CM are presented. In the Methods section, the experiments, datasets and algorithms are described. In the Results section, the fitting as well as the performance evaluation of the algorithms are presented. Finally, in the Discussion section we comment on the results and highlight the major contributions of the paper.

## 2 | RELATED WORK

### 2.1 | Choice models

Chaptini proposed the first application of discrete choice models (CM) in the field of recommender systems (Chaptini, 2005). The goal was to provide personalized course recommendations for MIT students by means of a generalized mixed logit model fed with survey data. Some years later, Polydoropoulou and Lambrou continued the utilization of CM to recommend courses for seafarers and employees of the shipping industry (Polydoropoulou & Lambrou, 2012). The models were estimated with data gathered from questionnaires. The novelty was centred on the application of a Bayesian approach to update the estimated coefficients, which allows for different prior distributions to characterize individual preferences. A more sophisticated model, a multi-level nested multinomial logit one, was proposed by Jiang et al. in the quest of achieving both relevancy and diversity in the recommendation process (Jiang et al., 2014). Recently, the group of Ben-Akiva at MIT have explored the potential of CM in app-based recommender systems (Danaf et al., 2019), a setting in which the attributes of the alternatives may vary over time and therefore user's preferences need to be continuously updated. The updating method has been tested in the field of transportation with real choices of users collected in Switzerland.

In the field of tourism, our preliminary work has revealed the potential value of CM for gastronomic recommendations by comparing their performance with that of basic rating-based algorithms (Saavedra et al., 2016). Following a similar line of thought, Mottini and Leheritier analysed how CM can be used in the air travel industry to have a better understanding of flight choices (Mottini et al., 2018).

### 2.2 | CM and the quality of data

Two types of information can be gathered from users: explicit and implicit. Explicit ratings obtained directly from users are the most common information used in recommender systems. However, it has been pointed out that ratings have some important drawbacks (Claypool et al., 2001):

(1) the input of these opinions can alter normal pattern of browsing and reading, and (2) users may stop entering this data if they do not see the benefit. As a result, the ratings may not be a reliable source of data for the recommending process. The solution is the use of implicit information, that is, the type of data obtained using an indirect method without direct interrogation of the user. Actions such as mouse clicks, mouse movement, browsing times and user's choices can be recorded and used to derive user's interest and preferences (Peska & Vojtas, 2017). Choices are the key data proposed in this paper to feed both the CM and the TRSs.

### 2.3 | CM and patterns of human decision-making

Understanding how tourists make choices is crucial to develop efficient TRSs. On this regard, the ASPECT model comes handy as it describes six different decision-making strategies or patterns that humans may follow when facing a choice problem (Jameson et al., 2015): attribute-based, consequence-based, experience-based, socially-based, policy-based and trial-and-error-based choice. A relationship between these patterns and state-of-art recommendation algorithms can be found: content-based algorithms follow the ideas behind the attribute-based pattern while collaborative-based algorithms could be related with the socially-based one. The CM utilized in this work may be considered as a formal implementation of the attribute-based choice pattern.

## 3 | BACKGROUND ON CHOICE MODELS

### 3.1 | Recommendation as a choice problem

The recommendation problem can be approached in different ways by viewing it as the problem of predicting user's choices in any particular context. Under this perspective, the Rational Choice Theory can be considered the classic paradigm used to explain the choices made by rational agents (Sen, 1990). This theory assumes that any decision-maker will solve the decision-making problem by applying the following rule:

$$CR(A, \succeq) = \left\{ a' \in A \middle\| a' \succeq a, \ \forall a \in A \right\}, \tag{1}$$

where, *CR* represents 'choice rule', *A* is the choice set, the set of alternatives considered for the decision maker at the time of choice, and the $\succeq$ operator represents the relationship 'preferred to', or at least 'preferred'. The chosen alternative will therefore be that for which the decision-maker shows the greatest preference.

In order to build a predictive model on the basis of this rule, the researcher must replace the qualitative preference operator with a quantitative one that will enable numerical comparison between the benefit of each alternative. Utility theory comes to the rescue to solve this issue. One of its axioms states that it is possible to define a utility function such that,

$$a \succeq b \Leftrightarrow U(a) \geq U(b). \tag{2}$$

Therefore the choice rule in Equation (1) becomes:

$$CR(A, \geq) = \left\{ a' \in A \middle\| U(a') \geq U(a), \ \forall a \in A \right\}. \tag{3}$$

This rule is mathematically equivalent to the formulation of the general recommendation problem (Adomavicius & Tuzhilin, 2005), which is described in terms of a maximization problem:

$$\begin{gathered} a' = \underset{a \in A}{\arg\max}\, U(a) \Leftrightarrow \\ CR(A, \geq) = \left\{ a' \in A \middle\| U(a') \geq U(a), \ \forall a \in A \right\}. \end{gathered} \tag{4}$$

As the recommendation problem can be understood as a choice prediction problem, the powerful models and techniques developed in the latter field can be applied to generate recommendations.

### 3.2 | Choice models with random utility

The choice rule represents how decision-makers reach their decisions. However, in the real world, researchers do not have access to all of the information that decision-makers may handle to estimate the utilities. For a specific user $c_n$, the researcher only knows some attributes of the alternatives, labelled $x_j$, for all $a_j$ alternatives with $j \in \{1, \dots, J\}$. Therefore, the predicted utility can be decomposed as follows:

$$U(c_n, a_j) = V_{nj} + \epsilon_{nj}, \tag{5}$$

where, $V_{nj} = V(x_j)$ is the representative utility, which can be estimated on the basis of the observed factors, and $\epsilon_{nj}$ captures the unknown factors that cannot be observed by the researcher. This decomposition is fully general, as $\epsilon_{nj}$ is defined simply as the difference between the true utility $U_{nj}$ and the representative utility $V_{nj}$.

The uncertainty about $\epsilon_{nj}$ is handled as a random variable, and the researcher must make further assumptions about its probability distribution. The models derived under these assumptions are called random utility models (RUM) (McF, 1973).

From the researcher's perspective, the choice rule of Equation (3) for a decision-maker $c_n$, which is deterministic from the decision-maker's perspective, becomes probabilistic in the following way:

$$CR(A, \geq) = \left\{a_i \in A \middle\| \mathbb{P}_{ni} \geq \mathbb{P}_{nj}, \forall a_j \in A\right\}, \tag{6}$$

and the probability $\mathbb{P}_{ni}$ is estimated by considering the decomposition formulated in Equation (4):

$$\mathbb{P}_{ni}\left(U(c_n, a_i) > U(c_n, a_j) \text{for all } j \neq i\right) = \mathbb{P}_{ni}\left(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \text{for all } j \neq i\right). \tag{7}$$

If the joint density of $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nJ})$ is denoted by $f$, the cumulative probability can be rewritten as follows:

$$\mathbb{P}_{ni} = \int_{\epsilon} \mathbb{I}\left(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj} \text{ for all } j \neq i\right) f(\epsilon_n) d\epsilon_n, \tag{8}$$

where, $\mathbb{I}$ is the indicator function, equalling 1 when the term in parentheses is true and 0 otherwise.

### 3.3 | Standard and mixed logit models

Different models are derived depending on the density chosen, that is, depending on the evidence or assumptions about the distribution of the unobserved portion of utility. The simplest and most widely adopted choice model is the standard logit model (McF, 1973), which is obtained under the assumption that each unobserved portion of utility $\epsilon_{nj}$ is distributed independently and identically. In this case, $f$ denotes the density for Gumbel distribution and the integral 8 takes a closed form with the following solution:

$$\mathbb{P}_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}. \tag{9}$$

This model estimates the probability $\mathbb{P}_{ni}$ as the ratio between the relevancy of the item $a_i$ for user $c_n$, estimated by the $e^{V_{ni}}$ term, and the aggregated relevancy of all items $a_j$ in the choice set. This set is the collection of items that the user considers/analyzes at the time of choice. Typically it is a reduced number of items that were filtered by the user by considering different constraints (price, distance, knowledge of the user, etc.).

The values of the probability $\mathbb{P}_{ni}$ depend on the representative utilities. As $V_{ni}$ increases, reflecting a higher match between the observed attributes of the alternative and the preferences of the decision-maker, with $V_{nj}$ for all $j \neq i$ held constant, $\mathbb{P}_{ni}$ approaches the value one. $\mathbb{P}_{ni}$ approaches zero when $V_{ni}$ decreases, as the exponential in the numerator approaches zero as $V_{ni}$ approaches $-\infty$.

The representative utility is usually specified as linear in the set of alternative attributes: $V_{nj} = \beta_{nj} \cdot x_j$, where $x_j$ is a vector including, as before, the observed attribute's values of the alternative $a_j$, and $\beta_{nj}$ denotes the model coefficients vector describing the preferences of decision-maker $c_n$ for the attributes of the alternatives $a_j$. The preferences $\beta_{nj}$ (model coefficients) are estimated by fitting Equation (9) to a data set of choices. The choice set must verify three properties. It must be finite, exhaustive (the decision-maker always chooses one of the alternatives) and mutually exclusive (the choice of one alternative necessarily implies not choosing any of the other ones).

The standard logit model cannot represent differences in tastes that are not related to observed characteristics (Train, 2009). Therefore, if taste variation is modelled as partly random, a logit model with random parameters should be considered instead. Thus, $\beta$ is now a vector of random coefficients that vary across decision-makers in the population with density $g$. This density is a function of parameters $\theta$ that represent, in the Gaussian case, the mean and covariance of the random coefficient in the population. The choice probabilities can be written as follows:

$$\mathbb{P}_{ni} = \int \left( \frac{e^{V_{ni}(\beta)}}{\sum_j e^{V_{nj}(\beta)}} \right) g(\beta|\theta) d\beta. \tag{10}$$

As the previous integral does not adopt a closed form, it must be evaluated numerically. Once the researcher specifies a distribution $g$ for the coefficients, the parameters $\theta$ maximizing the simulated log-likelihood must be estimated through simulation. The $R$ draws of the coefficients are then taken from $g$ and the logit probabilities are computed for each draw. The unconditional probability in Equation (10), which is the expected value of the conditional probabilities, is estimated as the average of the $R$ probabilities determined previously.

### 3.4 | Required data

In order to fit a choice-based model, we need a sufficient number of choices taken by the decision-maker. For each choice, the following data is required:

- The vector $x_i$ for the chosen alternative.
- The vectors $x_j$ for all alternatives $a_j$ in the choice set.

## 4 | METHODS

The methods were chosen to compare the performance of CM against rating-based models and popular ensemble strategies. The analysis was carried out with a gastronomic data set generated in an ecological experiment in the tourism domain. The design is described in Section 4.1 and the data set in Section 4.2. The details of the CM considered in this study are presented in Section 4.3. The baseline algorithms (rating-based and ensembles) chosen to compare our models are introduced in Section 4.4. The evaluation criteria used to estimate the performance of each algorithm are included in Section 4.5. Finally, software and implementation details are provided in Section 4.6.

### 4.1 | Experiment

We designed an ecological experiment under the scope of the RECTUR project. The chosen setting was the fourth edition (in 2011) of the Santiago(é)Tapas contest, a gastronomic event that takes place every year in the city of Santiago de Compostela. For the event, 56 local restaurants proposed and elaborated up to three tapas that were sold at a fixed price. A total of 5517 participants, including local, Spanish and international users, tasted the available tapas over a period of 2 weeks. A TapasPassport was made available to all participants and included the following official information: (i) the contest guidelines, (ii) restaurant location, and (iii) the tapas offered at each restaurant. After consuming the tapas, participants evaluated their experience by providing a vote with two ratings (Figure 1): (i) a rating for the tapas, and (ii) a rating of the overall experience (service, place atmosphere, etc.).

It is important to point out that the experiment was carried out in a real setting rather than a laboratory setting. Thus, the restaurants were free to offer whatever type of tapas they wished, and the participants made their own decisions about which tapas to try. It can therefore be assumed that the data set will include some sampling bias that may have some impact on the model predictions.

### 4.2 | RECTUR datasets

The data collected in the experiment were used to build two datasets: the choice and the rating dataset. The choice dataset consisted on choice observations, where each observation included the vector $x_i$ and the vectors $x_j$ containing the attribute values of the chosen tapa $a_i$ as well as the tapas $a_j$ of the choice set. To describe a tapa, the following attributes were considered (Table 1): type and character. Traditional tapas were


**FIGURE 1** RECTUR experiment. Images of votes, participating locals, and TapasPassport. The experiment was carried out in Santiago de Compostela during the celebration of a real contest of tapas

**TABLE 1** Tapa and tapa-user attributes

| Tapa attribute | Values |
|---|---|
| ID | $t1$ to $tn$ |
| Type | Cheese, Egg, Fish, Meat, Vegetable, Shellfish, Sweet, and Other |
| Character | Traditional or daring |
| **Tapa-user attribute** | **Values** |
| Rating | 0–5 |

created following well-known, popular recipes, while daring tapas were new and creative. In terms of data preparation, *type* and *character* attributes were transformed into eight dichotomous or binary variables associated with each value. Suppose the observation of a decision-maker located in the old area of the city choosing tapa $t100$, which is of a meat *type* and has a daring *character*. In this case, the chosen tapa was codified as follows: (1) 'meat' variable set to 1, (2) all other *type* variables set to 0, (3) 'daring' variable set to 1, and (4) 'traditional' variable set to 0. The choice dataset codified this way was used to fit the choice-based models.

On the other hand, the rating dataset stored a collection of ratings, an attribute of the tapa-user interaction, to gather the user's satisfaction with the tapa. The rating dataset was applied to train the baseline models.

## 4.3 | Choice models: standard and mixed logit models

The standard logit model as well as the mixed logit model, assuming Gaussian distribution on the coefficients, were chosen as basic representatives of the family of random utility choice-based models. Application of the mixed logit model was justified as we found evidence of taste variations among decision-makers on the basis of both personal and contextual factors (Ismoilov, 2017).

Although a large number of users tasted more than one tapa, the number of choices per user were not enough to fit a choice model per user. Constrained by this limitation, we decided to define three choice problems, each one corresponding to each area of the city. Each choice problem therefore aggregated the observations of all choices on each area of the city and assumed an unique choice set for all users in that area. Both the standard and mixed logit models were estimated for the three problems.

## 4.4 | Baselines

The following types of baseline models were chosen: (1) basic rating-based collaborative filtering algorithms, (2) advanced rating-based collaborative algorithm, (3) single decision-trees, and (4) tree-based ensemble strategies. The choice of tree-based from among other types of ensembles is explained by the fact that trees produce meaningful predictions (Ali et al., 2015; Quinlan, 1986), and they thus become a natural alternative to choice-based models to overcome the interpretability problem. Moreover, tree-based methods have proved useful for building recommender systems in different areas outperforming other approaches (Utku et al., 2015). However, the accuracy of prediction may suffer from both the size of the available data set (Bar et al., 2013; Ghimire et al., 2012) and the number of features (Lavanya & Rani, 2012). Our previous studies have demonstrated the superior performance of tree-based ensembles relative to single decision-trees, as well as their dependency on the number of available features (Almomani et al., 2017).

### 4.4.1 | Basic rating-based collaborative filtering (CF)

Two basic rating-based CF models were used: user-based collaborative filtering (CF-UB) and matrix factorization (CF-MF). CF-UB assumes that individuals with similar preferences will rate items in a similar way. Thus, missing ratings for a specific user $c_n$ can be predicted by finding a neighbourhood $N(n)$ of similar users and aggregating their ratings to calculate the corresponding prediction. The concept of similarity between users is used to define the neighbourhood given all users within a similarity threshold. In this study, the cosine similarity measure was considered, and $|N(n)|$ was set at 25. For an item $i$ and an individual $c_n$, the ratings predicted, $\widehat{r}_{ni}$, can be expressed as follows:

$$\widehat{r}_{ni} = \frac{1}{|N(n)|} \sum_{j \in N(n)} r_{ji}, \tag{11}$$

where, $||$ denotes the cardinality of $N(n)$.

CF-MF, on the other hand, characterizes both items and users by vectors of factors inferred from item rating patterns. For a given item $i$ and a user $c_n$, the vector $q_i$ measures the extent to which the item possesses those factors and the vector $p_n$, the extent of interest the user has in items that score highly on the corresponding factors. The dot product $q_i^T p_n$ captures the user's interest in the item's characteristics. This approximates user $c_n$'s rating of item $i$, $r_{ni}$, leading to the following estimate:

$$\widehat{r}_{ni} = q_i^T p_n. \tag{12}$$

Therefore, the challenge is to compute the mapping of each item and user to vectors $q_i$ and $p_n$. Here, singular value decomposition will be applied to factoring the user-item rating matrix, which may be sparse. In order to learn the factor vectors ($p_n$ and $q_i$), the regularized squared error on the set of known ratings is minimized:

$$\min_{q*,p*} \sum_{(u,\,i) \in K} \left(r_{ni} - q_i^T p_n\right)^2 + \lambda \left( \|q_i\|^2 + \|p_n\|^2 \right), \tag{13}$$

where, $K$ is the set of the ($c_n$, $i$) pairs for which $r_{ni}$ is known, $\|\|$ is the Euclidean norm and $\lambda$ denotes a constant controlling the extent of regularization. In this work, $\lambda = 1.5$.

### 4.4.2 | Advanced rating-based collaborative filtering (CF)

As a more complex model of this family, we resorted to CF-SVD++, an extension to CF-MF in which the effect of implicit information is included in the minimization rule. The difference here is that the prediction rule considers the fact of a user rating of an item as an additional indication of preference. Therefore, the vector representing the user's interest becomes (Koren, 2008):

$$p_n + |N(n)|^{-\frac{1}{2}} \sum_{j \in N(n)} y_j. \tag{14}$$

### 4.4.3 | Single decision trees

The idea of decision trees is to build a tree structure with nodes representing the features or attributes and leaves indicating the corresponding values of the attributes. The trees can be used for classification or for regression depending on the nature of the predicted outcome (Breiman et al., 1984). In this study, we used regression trees as the predicted variables (i.e., ratings) are numerical.

The tree is constructed through binary recursive partitioning, an iteration process that splits the features into branches. The process continues by splitting each partition into a minimum number of nodes. For the recursive binary splitting, both the splitting variable $X_i$ and a split point z are considered. The splitting at the split point is therefore described as follows:

$$R_1(i,z) = \{X | X_i \geq z\} \text{ and } R_2(i,z) = \{X | X_i < z\}. \tag{15}$$

A tree is formally described as follows:

$$T(X,\Theta) = \sum_{j=1}^{J} \gamma_j I(X \in R_j), \tag{16}$$

where a $\gamma_j$ parameter is assigned to each terminal node, and $\Theta = \{R_j, \gamma_j\}$. The prediction will be the mean of the outcome predictions (i.e., rating predictions in this study) in the region or terminal node. Figure 2 shows an example of the regression tree learnt for user number 1377 (u1377) and 44 tapas.

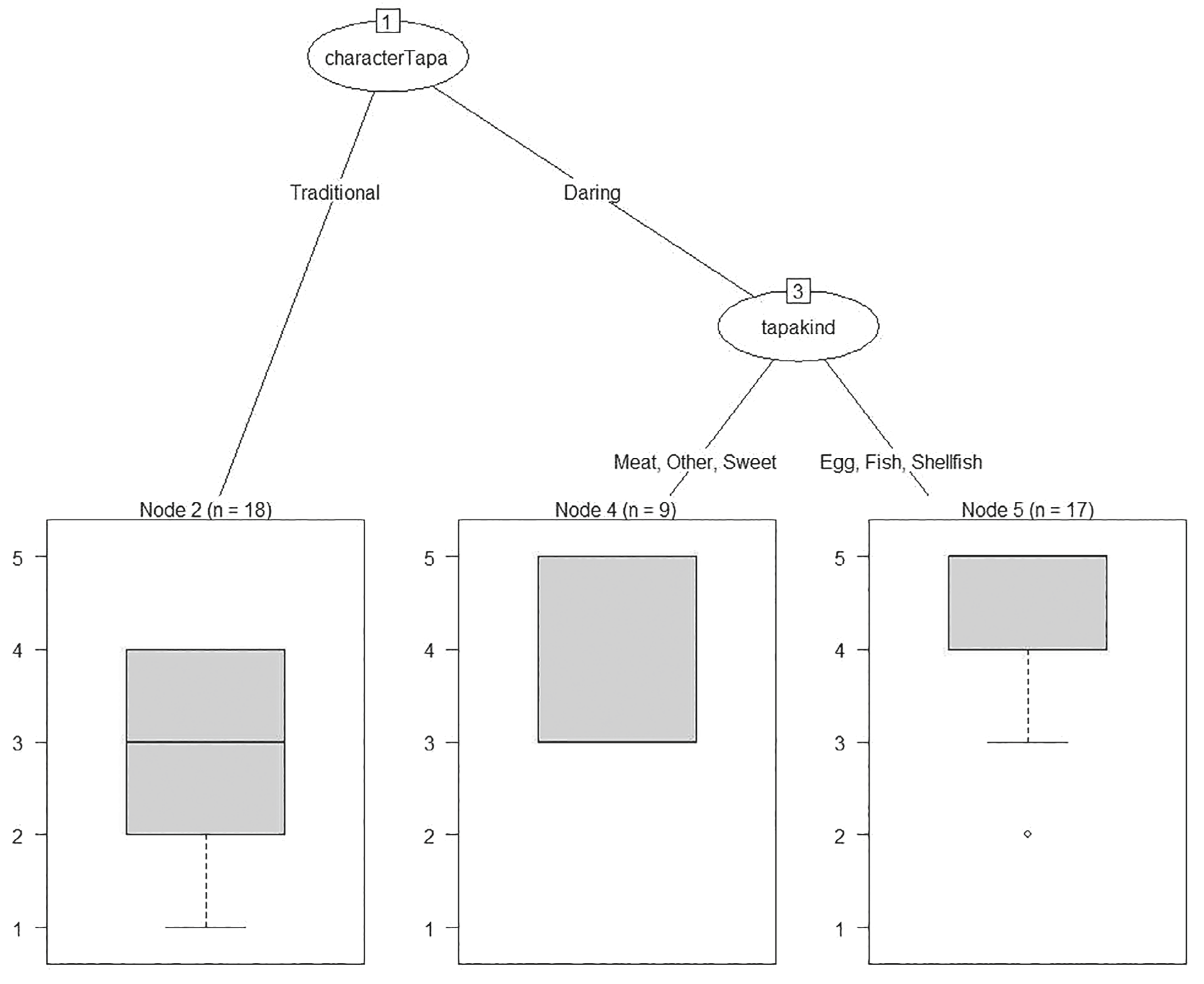


**FIGURE 2** Regression tree for user number 1377 learnt from 44 consumed tapas

### 4.4.4 | Tree-based ensemble strategies

Ensembles are aggregations of simple learners, such as trees, and the final prediction is estimated by combining the outcomes. The three ensemble methods used in this study are described below:

- *Boosting* builds a tree by using an iterative procedure. This means each tree depends and improves its performance on the basis of the prior trees. The prediction is estimated as follows:

$$f_m(x) = \sum_{m=1}^{M} T(X, \theta_m), \tag{17}$$

where, $M$ is the number of trees.

- *Bagging* builds different trees on $M$ different bootstrapped training data set. All trees are fully grown, indicating that a search over all features is carried out at each node in order to find the feature that best splits the data at that node. The final prediction is the average of each single tree estimation $\widehat{f}^{*b}(x)$:

$$\widehat{f}_{bag}(x) = \frac{1}{M}\sum_{m=1}^{M}\widehat{f}^{*b}(x). \tag{18}$$

- *Random Forests (RF)* is a particular case of Bagging. The main difference is that at each candidate split in the learning process, a random sample of the predictors or features is chosen among all the predictors or features. The goal is to build a large collection of uncorrelated trees. The prediction in a regression problem is estimated as follows:

$$\widehat{f}_{rf}(x) = \frac{1}{M}\sum_{m=1}^{M}\widehat{f}^{*b}(x). \tag{19}$$

## 4.5 | Evaluation

The performance of all models in the three areas of the city was analysed by applying random sub-sampling and leave-one-out cross validation to the RECTUR data set. For validation of random sub-sampling, 100 iterations were considered using 25% of randomly selected individuals for testing and the other 75% for training. For each decision-maker in the test data and for each recommendation method, prediction error measures were then estimated. The procedure for leave-one-out cross validation is similar, but the test set includes only one decision-maker per iteration.

Two metrics were applied in order to evaluate the performance of choice-based and rating-based algorithms: Precision and Discounted Cumulative Gain (DCG). For each tapas item included in the choice set, either its rating or its choice probability was predicted. Thereafter, the tapas were ranked and only the item with highest value was considered the predicted choice and therefore recommended (Top-1 scheme). The Precision measure was estimated as the fraction of correct recommendations to total recommendations after comparing the predicted choices with the real ones (Salton & McGill, 1986):

$$Precision = \frac{\text{Correct recommendations}}{\text{Total Recommendations}}. \tag{20}$$

Discounted Cumulative Gain (DCG) was chosen as a measure of ranking quality to capture the distance between the true choice and the predicted choice (Järvelin & Kekäläinen, 2002). This is defined as follows:

$$\mathrm{DCG}_p = \sum_{i=1}^{p}\frac{\mathrm{rel}_i}{\log_2(i+1)}, \tag{21}$$

where, $p$ is a particular ranking position, and $rel_i$ is the weighted relevance at position $i$. In the present study, there was only one relevant tapas item, that is, the item chosen, and therefore $rel_i$ was set to 1 when the relevant item was at position $i$, and 0 otherwise.

### 4.6 | Software

We conduct the analyses in R, the free software environment for statistical computing. Specifically, we used the following packages: (1) the *mlogit* package to estimate the multinomial logit models (see Croissant, 2012 for further details), (2) the *caret* package to estimate the single decision-trees and the tree-based ensembles, and (3) the *recommenderlab* package to evaluate the rating-based baselines. In addition, some cross-validation functions were developed using the R language to analyse the model performance in terms of predictions.

## 5 | RESULTS

### 5.1 | Data description

The choice dataset characterized in Section 4.2 includes the choices of 5517 individuals regarding a set of 113 tapas available during the Santiago (é)Tapas contest. The three choice sets corresponding to the three locations of restaurants in the city are briefly described.

In the *new area of the city* 2030 users consumed 3888 tapas that were chosen from 37 alternatives: 18 of traditional character, and 19 of daring character. Figure 3 shows a histogram plotting the consumption of each tapa, a label indicating the main ingredient and the average rating per item. According to the data, $t22$ and $t61$ were the most popular choices and the average rating was greater than 3, which indicates a high level of satisfaction.

As for the *old area*, 3953 participants tasted 8948 tapas chosen from the set of 62 available tapas: 32 of traditional character, and 30 of daring character. Figures 4 and 5 show the total number of daring and traditional tapas that users consumed for the 62 possible choices, respectively. According to the data, $t101$ was the most common choice, while $t37$, $t103$ and $t102$ were rarely selected. With regard to the average, the lowest ratings correspond to $t21$ and $t94$, and the highest to $t11$ and $t99$.

Finally, in the *outlying area of the city*, the least popular area, 436 users consumed 743 tapas from 14 available choices: 3 of traditional nature and 11 of daring nature. As before, Figure 6 summarizes the data set for this area. According to this figure, $t44$, $t45$, $t104$ and $t105$ were the tapas

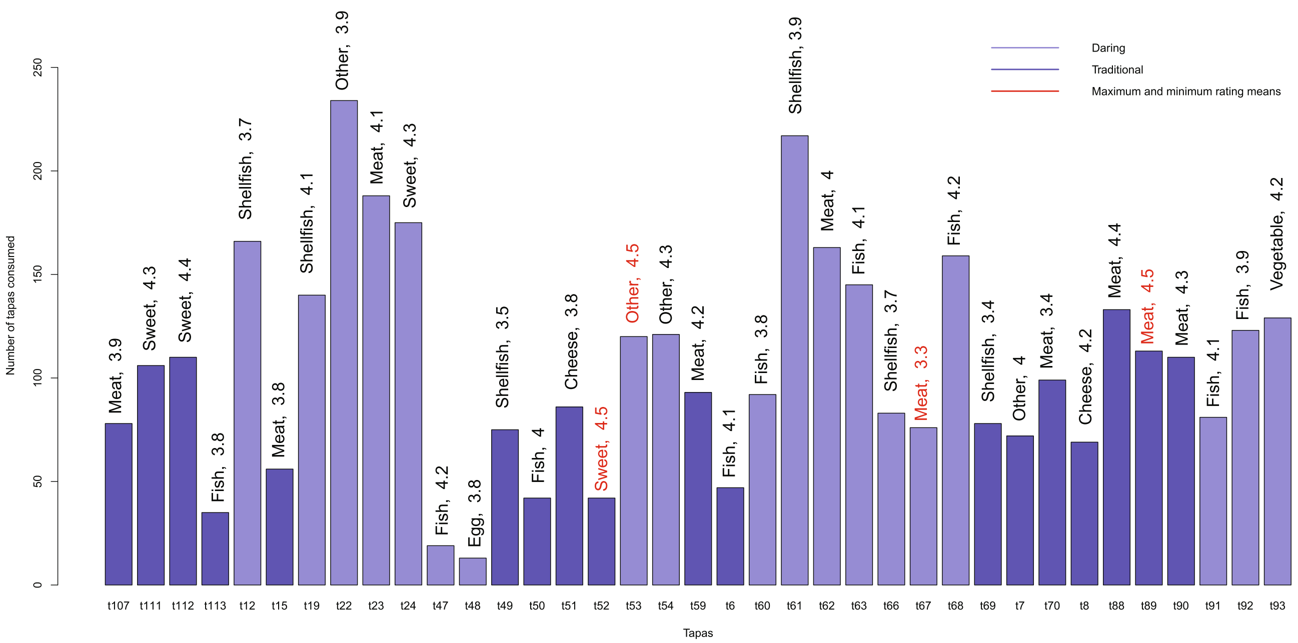


**FIGURE 3** Bar plot for number of different tapas consumed, main ingredient, and mean of users' ratings in the new area of the city

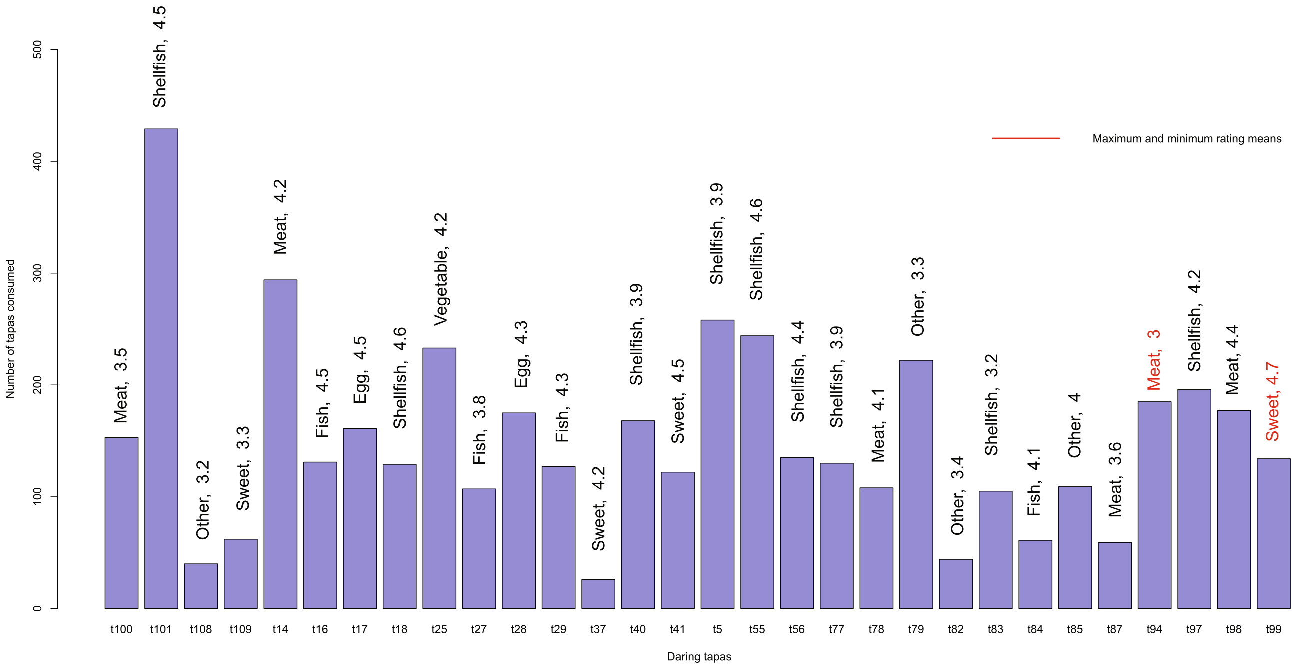


**FIGURE 4** Bar plot for number of different daring tapas consumed, main ingredient, and mean of users' ratings in the old area in the city

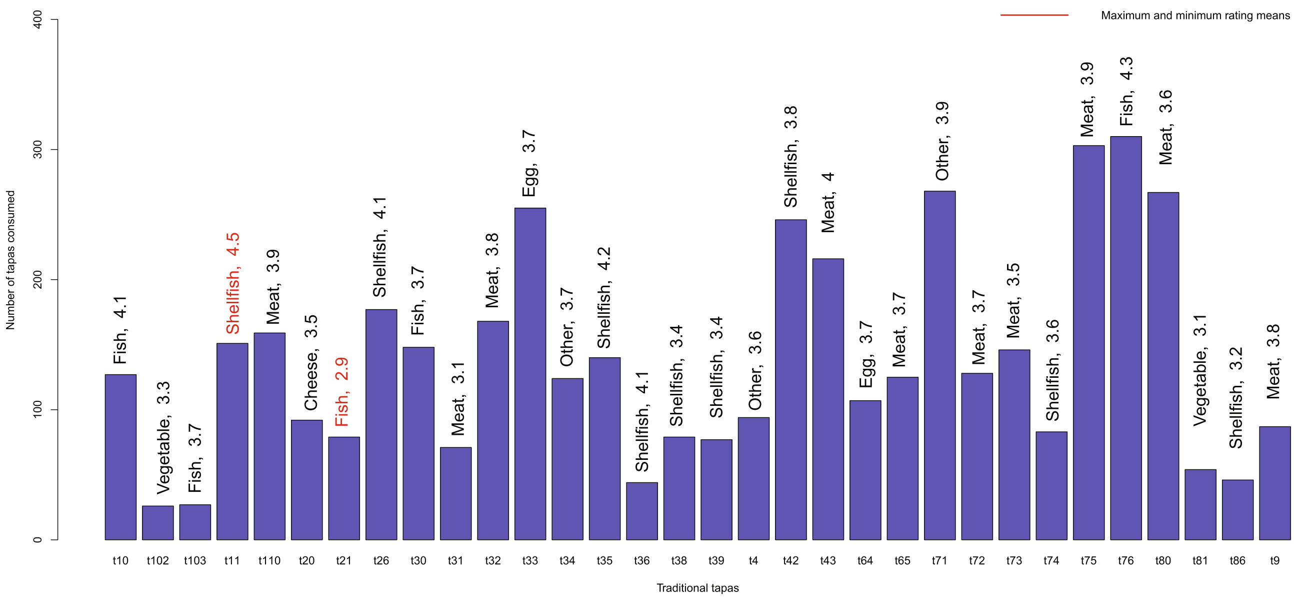


**FIGURE 5** Bar plot for number of different traditional tapas consumed, main ingredient, and mean of users' ratings in the old area

most frequently chosen, and *t*2 and *t*3 were rarely selected. In this case, the lowest and highest mean ratings correspond to tapas *t*58 and *t*44, respectively.

## 5.2 | Fitting of choice models

Both the standard and mixed logit models were fitted to the data for the three choice problems described in Section 4.3. For the mixed logit model, a Gaussian distribution of the coefficients was assumed, and the number of draws, *R*, was set to 100.

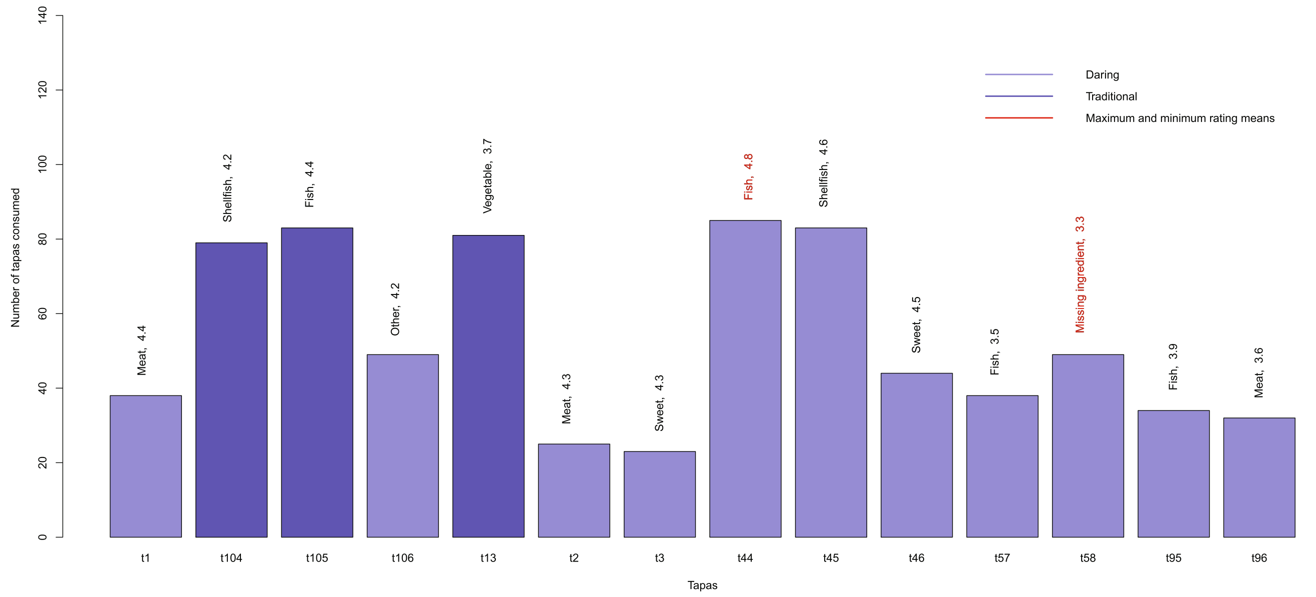


**FIGURE 6** Bar plot for number of different tapas consumed, main ingredient, and mean of users' ratings in the outlying area of the city

**TABLE 2** Estimation by maximum likelihood of the standard logit model coefficients for different areas of the city

| | New area | Old area | Outlying area |
|---|---|---|---|
| Cheese | −0.07 | **−0.25** | |
| Egg | **−2.48** | **0.31** | |
| Fish | **−0.46** | −0.02 | 0.14 |
| Meat | 0.06 | **0.28** | **−0.44** |
| Shellfish | −0.03 | **0.21** | **0.38** |
| Sweet | 0.07 | **−0.46** | **−0.38** |
| Vegetable | **−0.18** | **−0.17** | 0.26 |
| Traditional | **−0.62** | **−0.15** | **0.24** |
| Log-likelihood | −13,772 | −36,757 | −1913.8 |

*Note*: Significant coefficients are shown in bold.

The coefficients obtained for both models are shown in Tables 2 and 3. Most of these proved significant (in bold). For the mixed logit model (Table 3), only the mean estimations of Gaussian distributions are shown. In terms of preferences, the sign of coefficients represent the positive or negative preference of users for the tapas attribute. For instance, Table 2 shows that participants revealed a positive preference for egg, meat and shellfish tapas in the old area, but a negative one for egg and traditional tapas in the new area.

## 5.3 | Performance evaluation

According to the results shown in Section 5.2 the estimation of coefficients of the standard logit model is similar to the ones obtained for the mixed logit models. Therefore, only the first CM, the standard logit model, was evaluated and compared with the baseline algorithms.

Tables 4–6 show the evaluation results for the outlying, new and old areas of the city, respectively. The data shows that CM perform slightly better, in terms of Precision, than both CF and Ensemble algorithms, but quite similarly to the Single-Tree approach. However, in most cases, Precision is zero or close to zero, indicating that the predicted tapas item does not usually correspond to the one actually chosen. DCG, in turn, is more informative for analysing and comparing the performance of the different models. CM showed a superior performance in the Outlying area of the city, but Single-Tree and Ensemble algorithms provided better results in the other two areas.

**TABLE 3** Estimation of the means for mixed logit model coefficients assuming normal distribution for different areas of the city

| | New area | Old area | Outlying area |
|---|---|---|---|
| Cheese | −0.07 | −0.24 | |
| Egg | −**2.48** | **0.31** | |
| Fish | −**0.46** | −0.01 | 0.13 |
| Meat | −0.07 | **0.27** | −**0.67** |
| Shellfish | −0.03 | **0.21** | **0.37** |
| Sweet | −0.003 | −**0.46** | −**0.38** |
| Vegetable | −**0.18** | −**0.17** | 0.26 |
| Traditional | −**0.93** | −**0.09** | −0.01 |
| Log-likelihood | −13,631 | −36,680 | −1897.9 |

*Note*: Significant coefficients are shown in bold.

**TABLE 4** Outlying area of the city: cross validation predictions errors

| | R.CV | | LOO.CV | |
|---|---|---|---|---|
| Method | Precision | DCG | Precision | DCG |
| CHOICE | 0.11 | **0.35** | 0.12 | **0.36** |
| CF-UB | 0 | 0.29 | 0 | 0.29 |
| CF-MF | 0 | 0.29 | 0 | 0.29 |
| CF-SVD++ | 0.07 | 0.32 | 0.03 | 0.33 |
| Single-Tree | **0.13** | 0.33 | **0.16** | **0.36** |
| Ensemble-Boosting | 0 | 0.29 | 0 | 0.29 |
| Ensemble-Bagging | 0 | 0.29 | 0 | 0.29 |
| Ensembre-RF | 0 | 0.29 | 0 | 0.29 |

*Note*: Random and leave-one-out cross validation are denoted by R.CV and LOO.CV, respectively. In this area, the number of different tapas offered was 14 and DCG measures were estimated according to this ranking size.
Outlying area of the city: cross validation prediction errors. Best evaluation results are highlighted in bold.

**TABLE 5** New area of the city: cross validation predictions errors

| | R.CV | | LOO.CV | |
|---|---|---|---|---|
| Method | Precision | DCG | Precision | DCG |
| CHOICE | 0.05 | 0.27 | 0.05 | 0.25 |
| CF-UB | 0 | 0.21 | 0 | 0.21 |
| CF-MF | 0 | 0.21 | 0 | 0.21 |
| CF-SVD++ | 0.03 | 0.24 | 0.05 | 0.23 |
| Single-Tree | **0.14** | **0.41** | **0.11** | **0.40** |
| Ensemble-Boosting | 0 | 0.28 | 0 | 0.29 |
| Ensemble-Bagging | 0.07 | 0.40 | 0.05 | 0.39 |
| Ensembre-RF | 0 | 0.29 | 0 | 0.30 |

*Note*: Random and leave-one-out cross validation are denoted by R.CV and LOO.CV, respectively. In this area, the number of different tapas offered was 37 and DCG measures were estimated according to this ranking size.
New area of the city: cross validation prediction errors. Best evaluation results are highlighted in bold.

The ranking of the methods in terms of DCG in the R.CV validation is shown in Figure 7. In the case of equal DCG values, the ranking was based on Precision. Choice-models were ranked first in the outlying area (with only 14 alternative tapas) but the comparative performance was lower in the other two areas, in which the choice sets are larger. Ensembles show a somewhat opposite behaviour: they performed comparatively better when more choices, users and observations are available. Surprisingly, the Single-Tree method, including only one learner, provided

**TABLE 6** Old area: cross validation predictions errors

| Method | R.CV Precision | R.CV DCG | LOO.CV Precision | LOO.CV DCG |
|---|---|---|---|---|
| CHOICE | **0.02** | 0.21 | **0.01** | 0.21 |
| CF-UB | 0 | 0.18 | 0 | 0.18 |
| CF-MF | 0 | 0.18 | 0 | 0.18 |
| CF-SVD++ | 0.01 | 0.21 | **0.01** | 0.20 |
| Single-Tree | 0 | 0.21 | 0 | 0.21 |
| Ensemble-Boosting | 0 | **0.22** | 0 | **0.22** |
| Ensemble-Bagging | 0 | 0.20 | 0 | 0.20 |
| Ensembre-RF | 0 | 0.20 | 0 | 0.21 |

*Note*: Random and leave-one-out cross validation are denoted by R.CV and LOO.CV, respectively. In this area, the number of different tapas offered was 62 and DCG measures were estimated according to this ranking size.

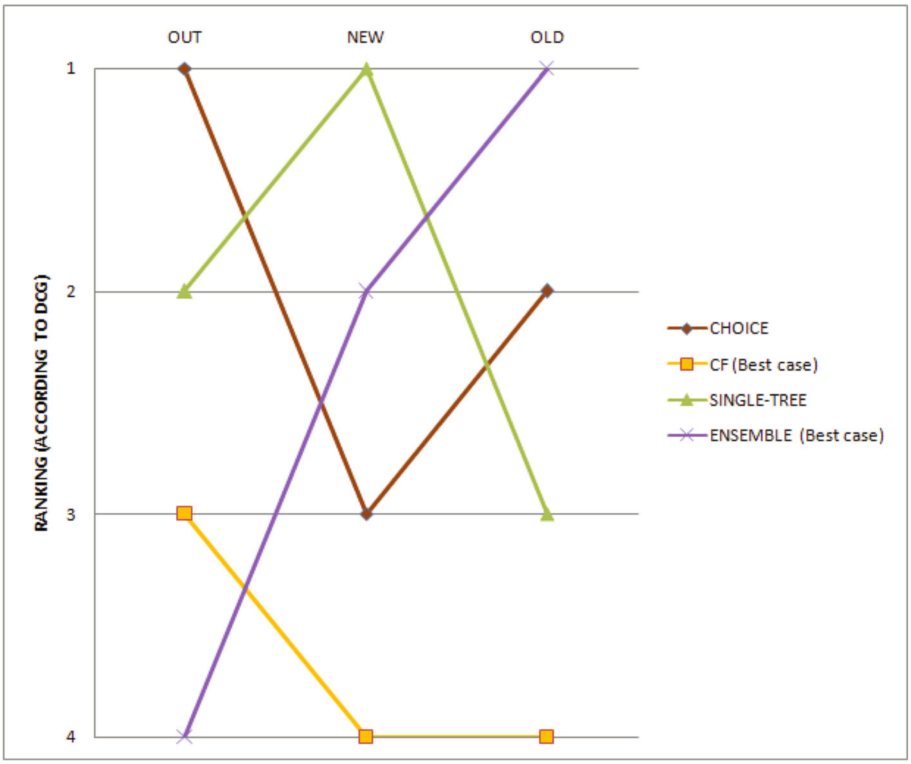


**FIGURE 7** Ranking of models according to DCG in each area of the city. The performance of models depends on the size of the data set. Choice models work better with fewer available users and fewer alternatives in the choice set (outlying area of the city), while the ensemble models perform better as the size of the data increases (old area)

competitive results in the outlying and new areas. As expected, however, for larger data sets (old part of the city) the performance was not as good as that of ensembles and CM. On the other hand, CF approaches are far from being competitive and occupied the lowest position, except for the outlying area of the city. Another interesting finding is that the performance of all models in terms of DCG was lower as long as the choice set, that is, the number of available tapas, increased from the outlying area to the old area. This may suggest that the prediction problem is probably more complex when the choice set increases and the choice becomes more difficult for the decision-maker.

## 6 | DISCUSSION

With the aim of providing algorithms for tourism recommender systems with both a theoretical understanding of tourist's motivations and a certain degree of transparency/interpretability, we have applied CM to generate recommendations of tapas. From this perspective, the recommendation problem is considered a problem of choice prediction rather than of rating prediction, the current paradigm in the field of recommender systems. The key elements of CM are as follows: (1) user's preferences are learnt from choices, (2) the choice set of each choice situation is treated as an important variable for both explaining and predicting future choices, and (3) unobserved factors affecting the decision-making process are captured through random variables.

### 6.1 | Benefits of choice models

CM offer a theoretical background to generate sound recommendations. The robustness of the approach is achieved by means of estimating preferences from a reliable source. In CM, preferences are learnt by fitting choice-based models with choice data. The underlying assumption is that user's choices are the result of the direct matching between the user's preferences and the item's attributes. So, by observing the choices and gathering the attribute's values of the alternatives in the choice sets, the unknown preferences can be discovered. On the other hand, rating-based models rely on user's ratings, which represent a post-experience satisfaction. This outcome mainly depends on the comparison between the real and the expected satisfaction. Preferences and tastes are thus out of this equation, meaning that ratings and preferences have no clear relationship.

Transparency and interpretability are another contribution of choice models. The estimated preferences, that is, model coefficients, provide a means of easily explaining why some items are more likely to be recommended than others. Moreover, this benefit does not mean a decrease in performance. The results shown in Tables 4–6, and summarized in Figure 7, suggest that choice models may be inferior to ensembles only in situations where larger data sets (old area) are available. The fact that ensembles show a superior performance in the case of larger datasets was expected as the *N* predictors of the ensemble could take advantage of sampling the data in a more efficient way.

### 6.2 | Cost of choice models

CM require more information than rating-based models. As described in Sections 3.4 and 4.2, CM need attribute's values as well as choice sets, while rating-based models only use ratings. The history of science shows that the improvement of the accuracy of a model has frequently come at the cost of increasing complexity. This often means that new variables and data need to be gathered in order to test the model. However, the cost of information gathering is alleviated by the fact that user's choices and their corresponding choice sets are collected in an implicit way. In our specific problem, choices are naturally recorded using the tapa votes and choice sets are automatically determined by the location of the restaurant. In a digital environment, this process would be even easier, as every user-environment interaction would be automatically recorded. So, by using choice data we remove the burden of explicitly interrogating decision-makers about their ratings. Moreover, the attributes and attribute values may be learnt automatically by means of Product Attribute Extraction (PAE) methods. If textual product descriptions, such as tags, product reviews, and so forth, were available, which happens quite often in e-commerce sites, the domain information can be gathered with no manual effort.

In short, both implicit information and PAE methods can naturally be applied in digital settings to reduce the cost of information gathering in a quite significant way. This will lower the information barrier to apply CM.

### 6.3 | Issues with our specific problem: bias and scalability

In our experiment the involved restaurants have different accessibility and popularity, which also have an impact on the tapas offered on each restaurant. This all means that our dataset is biased. It ends up on having some items that a priori seems to have higher or lower probability of being chosen. As a result, we cannot guarantee an orthogonal design to ensure that each tapa and restaurant class is in the choice set available to be chosen exactly the same number of times. A possible solution to this problem might be based on embedding the proposed choice models in a Bayes probabilistic scheme, so the sampling bias could be managed as prior probabilities. This is a very interesting issue and we will face the problem in our future work.

Another concern is about the scalability of CM in terms of available alternatives. In order to understand this issue, two related but different concepts need to be defined: (1) the user's choice set, and (2) the recommendation set. The former is the reduced set of items that the user is considering/analysing at the time of choice. The later is the set containing all the items that are available at the time of the recommendation. The first set is usually a small subset of the second set. Now, two computational tasks need to be carried out:

- Fitting/Training task. For model fitting we need to know the user's choice set. The size of the choice sets determine the total elements in the summation of Equation (9). This summation is just a normalization factor to ensure that the model provides a probability value. In our tapa's problem the models deal with quite large choice sets. This is so as we assumed the user may consider to taste any tapa available on the region she would be located. Therefore, all the tapas available on each of the three areas of the city should belong to the choice set of that area. In fact, by making this assumption, the choice set coincides with the recommendation set. In general, the choice set should be much smaller than the ones used in our paper, so we believe that the size of the choice set will not be an issue when applied to other problems.

- Prediction/Recommendation task. At this stage, the recommendation set is used to estimate the choice probability per each alternative. In this stage, the number of alternatives in the recommendation set is not a real problem, as it only impacts on the computational cost of running $N$ probability estimations. In a real e-commerce scenario where timely recommendations are required and $N$ may be extremely large, various solutions could be implemented, such as reducing the recommendation set (viewed alternatives, purchased alternatives by similar users, etc.) and/or periodically updating the probability estimations beforehand.

## 6.4 | Conclusion

CM seem to provide an optimal trade-off between theoretical soundness, interpretability and performance in the field of tourism recommender systems. For future work, we plan to go a step further in terms of uncovering motivational drivers of behaviour. For instance, we wonder about the effect of cultural factors in the mindset of tourists and how they affect their choices. We are also interested on analysing different choice problems in the field of tourism in order to prove the efficiency of CM in a broader scope.

In summary, CM pave the way to the application of sound decision-making models in the field of tourism recommender systems, and open the door of incorporating new and powerful motivational factors to develop more efficient algorithms.

### ACKNOWLEDGEMENTS

This research was sponsored by EMALCSA/Coruña Smart City under grant CSC-14-13, the Ministry of Science and Innovation of Spain under grant TIN2014-56633-C3-1-R, the Ministry of Economy and Competitiveness of Spain under grant MTM2013-41383P, the Consellería de Cultura, Educación e Ordenación Universitaria (accreditation 2016-2019, ED431G/08), and the European Regional Development Fund (ERDF).

### CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734–749.

Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender systems handbook* (pp. 217–253). Springer.

Ali, S., Tirumala, S. S., & Sarrafzadeh, A. (2015). Ensemble learning methods for decision making: Status and future prospects. In *International conference on machine learning and cybernetics (ICMLC)* (pp. 211–216). IEEE.

Almomani, A., Saavedra, P., & Sánchez, E. (2017). Ensembles of decision trees for recommending touristic items. In *International work-conference on the interplay between natural and artificial computation* (pp. 510–519). Springer.

Bar, A., Rokach, L., Shani, G., Shapira, B., & Schclar, A. (2013). Improving simple collaborative filtering models using ensemble methods. In *International workshop on multiple classifier systems* (pp. 1–12). Springer.

Borràs, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, *41*(16), 7370–7389.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth and Brooks.

Burke, R., Felfernig, A., & Goker, M. (2011). Recommender systems: An overview. *AI Magazine*, *32*(3), 13–18.

Chaptini, B. H. (2005). *Use of discrete choice models with recommender systems* [PhD thesis]. Massachusetts Institute of Technology.

Claypool, M., Le, P., Wased, M., & Brown, D. (2001). Implicit interest indicators. In *Proceedings of the 6th international conference on intelligent user interfaces* (pp. 33–40).

Croissant, Y. (2012). *Estimation of multinomial logit models in r: The mlogit packages*. R package version 02-2.

Crompton, J. L. (1979). Motivations for pleasure vacation. *Annals of Tourism Research*, *6*(4), 408–424.

Danaf, M., Becker, F., Song, X., Atasoy, B., & Ben-Akiva, M. (2019). Online discrete choice models: Applications in personalized recommendations. *Decision Support Systems*, *119*, 35–45.

Dann, G. (1976). The holiday was simply fantastic. *The Tourist Review*, *31*(3), 19–23.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer-Verlag.

Ghimire, B., Rogan, J., Galiano, V. R., Panday, P., & Neeti, N. (2012). An evaluation of bagging, boosting, and random forests for land-cover classification in cape cod, Massachusetts, USA. *GIScience & Remote Sensing*, *49*(5), 623–643.

Hamid, R. A., Albahri, A. S., Alwan, J. K., Al-Qaysi, Z., Albahri, O. S., Zaidan, A., Alnoor, A., Alamoodi, A. H., & Zaidan, B. (2021). How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management. *Computer Science Review*, *39*(100), 337.

Harman, D. (1995). Overview of the 3rd text retrieval conference (trec3). In *The 3rd text REtrieval conference (TREC-3)* (pp. 1–20). Department of Commerce, National Institute of Standards and Technology.

Ismoilov, J. (2017). *Stated and revealed preferences on gastronomic tourism in Santiago de Compostela* [PhD thesis]. University of Santiago de Compostela.

Jameson, A., Willemsen, M. C., Felfernig, A., de Gemmis, M., Lops, P., Semeraro, G., & Chen, L. (2015). Human decision making and recommender systems. In *Recommender systems handbook* (pp. 611–648). Springer.

Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, *20*(4), 422–446.

Jiang, H., Qi, X., & Sun, H. (2014). Choice-based recommender systems: A unified approach to achieving relevancy and diversity. *Operations Research*, *62*(5), 973–993.

Koren, Y. (2008). Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 426–434). ACM.

Kzaz, L., Dakhchoune, D., Dahab, D., Park, D., Kim, H., Carrer-Neto, W., Hernández-Alcaraz, M., Valencia-García, R., Meehan, K., & Lunney, T. (2018). Tourism recommender systems: An overview of recommendation approaches. *International Journal of Computers and Applications*, *180*(20), 9–13.

Lavanya, D., & Rani, K. U. (2012). Ensemble decision making system for breast cancer data. *International Journal of Computer Applications*, *51*(17), 19–23.

Le, Q. T., & Pishva, D. (2016). An innovative tour recommendation system for tourists in Japan. In *2016 18th international conference on advanced communication technology (ICACT)* (pp. 717–729). IEEE.

McF (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*. Academic Press.

Mottini, A., Lhéritier, A., Acuna-Agost, R., & Zuluaga, M. A. (2018). *Understanding customer choices to improve recommendations in the air travel industry* (pp. 28–32). RecTour@ RecSys.

Peska, L., & Vojtas, P. (2017). Using implicit preference relations to improve recommender systems. *Journal on Data Semantics*, *6*(1), 15–30.

Pestana, M. H., Parreira, A., & Moutinho, L. (2020). Motivations, emotions and satisfaction: The keys to a tourism destination choice. *Journal of Destination Marketing & Management*, *16*(100), 332.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, *6*(3), 21–45.

Polydoropoulou, A., & Lambrou, M. A. (2012). Development of an e-learning recommender system using discrete choice models and bayesian theory: A pilot case in the shipping industry. In *Security Enhanced Applications for Information Systems*. IntechOpen.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on computer supported cooperative work* (pp. 175–186).

Saavedra, P., Barreiro, P., Duran, R., Crujeiras, R., Loureiro, M., & Vila, E. S. (2016). *Choice-based recommender systems* (pp. 38–46). RecTour@ RecSys.

Salton, G., & McGill, M. J. (1986). *Introduction to modern information retrieval*. McGraw-Hill, Inc.

Santos, F., Almeida, A., Martins, C., Gonçalves, R., & Martins, J. (2019). Using poi functionality and accessibility levels for delivering personalized tourism recommendations. *Computers, Environment and Urban Systems*, *77*(101), 173.

Sen, A. (1990). Rational behaviour. In *Utility and probability* (pp. 198–216). Springer.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Utku, A., Karacan, H. U., Yildiz, O., & Akcayol, M. A. (2015). Implementation of a new recommendation system based on decision tree using implicit relevance feedback. *JSW*, *10*(12), 1367–1374.

Wan, L., Hong, Y., Huang, Z., Peng, X., & Li, R. (2018). A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks. *International Journal of Geographical Information Science*, *32*(11), 2225–2246.


## AUTHOR BIOGRAPHIES

**Ameed Almomani** was born in Irbid, Jordan in 1980. He graduated from Yarmouk University (YU) in Jordan in 2002 in the field of computer science. After that, he studied a master's degree in computer information systems in Jordan as well, and graduated in 2006. He received a PhD in Artificial Intelligence from CiTIUS, one of the research centers of Santiago de Compostela University (USC). He worked in the Ministry of Education in Jordan and then as a lecturer at several universities in the Kingdom of Saudi Arabia. Nowadays, he is working with Artgro company as a Datascientist in the USA. His main research interests are in Artificial Intelligence, Recommender Systems, Machine learning, Choice modeling, and decision making.

**Paula Saavedra** (Santiago de Compostela, 1984) is an Assistant Professor at the Department of Statistics, Mathematical Analysis and Optimization (USC). She obtained her PhD in Statistics and Operations Research in March 2015, with a thesis entitled "Nonparametric data-driven methods for set estimation" (supervised by Prof. Wenceslao González-Manteiga [USC] and Alberto Rodríguez-Casal [USC]). Her main research is focused on nonparametric statistics methods for set estimation.

**Pablo Barreiro** received his Master Degree in Information Technologies in 2014. He worked as a research assistant at the University of Santiago de Compostela (Spain), in the field of choice-based recommendation algorithms. Nowadays, he left the academic world to join the software development industry, where he is working as a Senior Software Engineer at Auctane.

**Roi Durán** received his Phd in Economics, with emphasis in environmental economics, University of Vigo. Currently, he works as a teacher of Secondary Education.

**Rosa Crujeiras** received her Ph.D. degree in Mathematics in 2007. She has been a postdoc researcher at Universitè Catholique de Louvain (Belgium) and currently she is an associate professor at Universidade de Santiago de Compostela (Spain). She is the Scientific Director of the

Galician Centre for Mathematical Research and Technology (CITMAga). Her main research interest are focused on nonparametric statistical methods for dependent and complex data.

**María Loureiro** is a Professor of Economics at University of Santiago de Compostela, Spain; PhD from Washington State University, USA. Researcher included in the Highly Cited Edition 2012, and on the top 2% most influential researchers according to Ioannidis JPA, Boyack KW, Baas J (2021) "Updated science-wide author databases of standardized citation indicators." PLoS Biol 18(10): e3000918. https://doi.org/10.1371/journal.pbio.3000918. See recent data at https://elsevier.digitalcommonsdata.com/datasets/btchxktzyw/3.

**Eduardo Sánchez** is an Associate Professor at the University of Santiago de Compostela. He obtained a Ms in Neuroscience at the International University of Andalucía, and a Ms in Computer Science and Software Engineering at the University of Southern California. He received his PhD degree in Physics in 2001. His research is focused in the fields of decision-making and recommender systems, developing predictive models in the domains of tourism and culture, and computational neuroscience, working with models of the visual system.

**How to cite this article:** Almomani, A., Saavedra, P., Barreiro, P., Durán, R., Crujeiras, R., Loureiro, M., & Sánchez, E. (2023). Application of choice models in tourism recommender systems. *Expert Systems*, *40*(3), e13177. https://doi.org/10.1111/exsy.13177