



HAL
open science

A machine learning approach for analyzing Free JAR data

Alexiane Luc, Sébastien Lê, Mathilde Philippe, El Mostafa Qannari, Evelyne Vigneau

► **To cite this version:**

Alexiane Luc, Sébastien Lê, Mathilde Philippe, El Mostafa Qannari, Evelyne Vigneau. A machine learning approach for analyzing Free JAR data. Food Quality and Preference, 2022, 99, pp.article n° 104581. 10.1016/j.foodqual.2022.104581 . hal-03675589

HAL Id: hal-03675589

<https://hal.science/hal-03675589v1>

Submitted on 22 Jul 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **A machine learning approach for analyzing Free JAR data**

2

3 *Alexiane LUC^{a,b,c}, Sébastien LE^c, Mathilde PHILIPPE^b, El Mostafa QANNARI^a, Evelyne VIGNEAU^a*

4

5 ^a ONIRIS, INRAE, StatSC, Nantes, France

6 ^b Strategir, Bordeaux, France

7 ^c Institut Agro, Univ Rennes, CNRS, IRMAR, Rennes, France

8

9 **Corresponding author:**

10 Alexiane LUC

11 aluc@strategir.com

12

13 **ABSTRACT:**

14 This paper presents a methodology based on a machine learning approach for the processing of Free
15 JAR data. The development of this new analysis method is grounded on the fact that Free JAR data
16 can be seen as labeled textual data, since a hedonic category is associated to each Free JAR
17 comment. In particular, this two-step methodology aims to take advantage of the link between these
18 two pieces of information, to disclose the products' assets and weaknesses. The first step consists of
19 assigning valency scores to the Free JAR comments, considering their link to the hedonic
20 categorization provided by the consumers. This is achieved by setting up a classifier based on
21 Random Forest. The second step consists of highlighting the attributes that characterize the product
22 under study based on the concept of interpretability. The interest of the proposed procedure is
23 illustrated through a case study pertaining to cheeses.

24

25 **KEYWORDS:** Free JAR; Labeled textual data; Valency score; Interpretability; Preference mapping;
26 Cheese

27

1. Introduction

In a product development context, researchers aim to clearly identify the products' assets and weaknesses perceived by the consumers. Several approaches allow us to obtain such information. Among these procedures, we single out Just-About-Right (JAR) (Rothman & Parker, 2009) and Check-All-That-Apply (CATA) (Ares et al., 2010; Ares & Jaeger, 2015). These procedures are based on predefined lists of attributes. By contrast, in alternative procedures, it is advocated to let the consumers freely express their perception of the products. For instance, this can be achieved by means of open-ended questions, which allow the respondents to answer in text format without imposing a limited set of options (ten Kleij & Musters, 2003). Clearly, the gain from such alternatives is to collect spontaneous assessments from the consumers. However, since the answers from the various respondents have different patterns, it is not easy to handle and analyze this kind of data.

In psychology, the term *valence* is used to identify the intrinsically pleasant or unpleasant quality of a stimulus (Lewin, 1935). An asset can be defined as a word or a set of words associated with a positive valence. A weakness can be defined as a word or a set of words associated with a negative valence. The difficulty encountered in analyzing textual data is to understand the valence of words, be it positive or negative, when they are associated with other words in the same sentence. Sentiment analysis was developed to address this issue (Mohammad, 2016). This type of analysis is relatively new in sensory evaluation and several related studies have been carried out during the past ten years to better pinpoint the consumers' perceptions and determine their emotions towards the products (King et al., 2013; Gutjar et al., 2015; Luc et al., 2020; Visalli et al., 2020). However, sentiment analysis entails referring to dictionaries or labeling the stimuli according to their positive or negative features and, as long as these dictionaries or labels are set up by a human being, there is a risk of bias.

The Free JAR procedure (Luc et al., 2020) yields textual data with the aim of disclosing the products' assets and weaknesses. For this effect, the respondents are allowed to freely express their assessments but they are encouraged to use such terms as "too much", "not enough" and "just about right (JAR)". Prior to this assessment, they are instructed to sort the products in three hedonic categories: "I like very much", "I like moderately" and "I don't like". This categorization task serves, among others, as a nudge to prompt the respondents to focus on the assets and weaknesses of the products being assessed.

In a first paper introducing the Free JAR procedure (Luc et al., 2020), we proposed a procedure to compute a sentiment score per comment based on a lexical and dictionary-based approach of sentiment analysis (Mohammad, 2016). The present paper, which should be considered as a methodology-focused paper, aims to go further and presents a new way to obtain a score that represents the global valence expressed by a respondent toward a product. This is achieved by investigating the link between Free JAR comments and hedonic categories provided by the respondents themselves through machine learning methods. Furthermore, the paper aims to investigate an alternative approach for considering textual data, in a more flexible and less time-consuming way than specific and manual approaches, as in Luc et al. (2021). By comparison, the rationale behind the approach of analysis outlined herein is to consider the data as labeled textual data.

The paper is organized as follows. In the first section, after sketching an outline of the Free JAR protocol, the different steps followed for the preprocessing of the textual data are explained. Then, we first focus on the analysis of the Free JAR data at the comment level. In particular, we aim to assign a valency score to each comment, reflecting its positive or negative nature. For this purpose, a procedure of analysis is developed. This consists in investigating the link between the hedonic categories and the Free JAR comments thanks to a Random Forest classifier. The implementation of

76 this classifier allows us to link the presence or absence of words to the corresponding hedonic
 77 category, in order to extract quantitative scores. Thereafter, an analysis at the level of the words that
 78 form the various comments is carried out in order to highlight the drivers of liking and disliking that
 79 were used to describe a given product. This latter analysis is based on a concept called
 80 interpretability (Miller, 2018). In a machine learning context, the interpretability reflects the extent
 81 to which a human being can understand the predictions of an algorithm. The higher the
 82 interpretability of a machine learning model, the easier it is to understand the rationale behind the
 83 prediction process. In fact, this understanding can help to learn more about the data (Heider, 1958;
 84 Lombrozo, 2006; Williams et al., 2013).

85 Figure 1 summarizes the different steps for the analysis of Free JAR data adopted in this paper.
 86 The general approach is illustrated based on a case study pertaining to cheese products. A discussion
 87 and concluding remarks are finally reported in light of the findings from the case study.

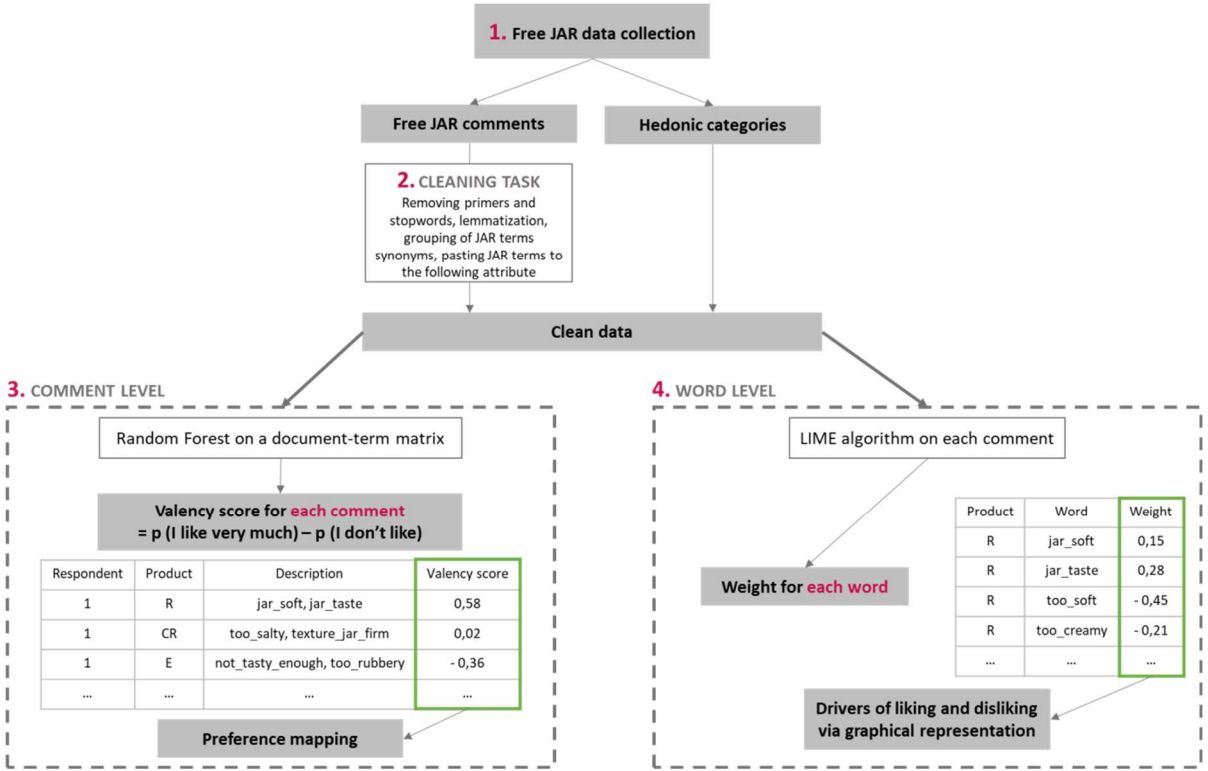


Figure 1: Overview of the different steps that will be followed in this paper for the analysis of Free JAR data.

<i>Comment</i>	<i>Hedonic category</i>	creamy	strong	sweet	too	...	just
----------------	-------------------------	--------	--------	-------	-----	-----	------

112 **2. Material and methods**
113

Respondent	Product	Hedonic category	Comment
1	A	1	This product is very sweet
1	B	3	This cheese is too creamy and too strong
1	C	3	This cheese is very characterful with its strong taste
2	A	2	This product doesn't have enough taste
2	B	1	This product is too creamy but just sweet enough
2	C	1	I like this sweet cheese

114 2.1. Free JAR protocol
115

116 The Free JAR procedure consists of two steps (Luc et al., 2020). Firstly, the respondents are
117 invited to sort the products into three hedonic categories, namely "I don't like", "I like moderately"
118 and "I like very much". Then, they are instructed to describe the products using their own attributes,
119 but they are prompted to use terms such as "not enough", "JAR" or "too much".

120 Free JAR data can be considered as labeled data in the sense that each Free JAR comment is
121 associated with a hedonic information, namely the hedonic category in which the respondent has
122 placed the product under consideration prior to the description step.

123 Let n be the number of respondents and p the number of products. The data resulting from a
124 Free JAR procedure can be stored as a table with 4 columns and $p * n$ rows, as illustrated in Table 1.
125 In this table, the hedonic categorizations are recoded as "1: I like very much", "2: I like moderately"
126 and "3: I don't like".

127
128 *Table 1: Raw data resulting from a Free JAR procedure.*
129

130 2.2. Data preprocessing
131

132 The raw Free JAR data can also be seen as sets of words associated with a hedonic category,
133 structured as a table with $p * n$ rows that correspond to the various comments, and as many
134 columns as there are words used by all the respondents. This data structure, commonly known as
135 document-term matrix (Borko, 1965), is exemplified in Table 2.
136

This product is very sweet	1	0	0	1	0		0
This cheese is too creamy and too strong	3	1	1	0	2		0
This cheese is very characterful with its strong taste	3	0	1	0	0		0
This product doesn't have enough taste	2	0	0	0	0		0
This product is too creamy but just sweet enough	1	1	0	1	1		1
I like this sweet cheese	1	0	0	1	0		0

137
138

Table 2: Data resulting from a Free JAR procedure, structured as sets of words.

139 To get the most out
140 of the Free JAR data, a

141 data preprocessing is recommended. The cleaning task is an important step when analyzing textual
142 data (Lebart et al., 1998; Piqueras-Fiszman, 2015).

143 In particular, the lemmatization of the words is one of the most common text preprocessing
144 techniques. This means that the various and different occurrences of a given word (singular/plural,
145 different conjugations of a verb, etc.) are reduced to their root origin, called a lemma, so that they
146 can be analyzed as a single item. This entails using a dictionary or a lexicon, as the one from
147 IRaMuTeQ© software (Ratinaud, 2014) that we used herein. The preprocessing of the data was also
148 achieved by removing the comment primers, such as "I like this product very much because..." or "I
149 don't like this product because...". Indeed, these primers were given to the respondents as examples
150 in order to help them better describe the products but are not relevant for the analysis. The so-called
151 stopwords such as "this", "you", "is" were filtered out from the comments because they were
152 considered as meaningless or irrelevant to the study (Bécue-Bertaut et al., 2008).

153 Then, each of the synonyms of the terms "too much", "not enough" and "just about right" were
154 grouped under the same term. This synonymous grouping has been achieved using a dictionary of
155 synonyms, which allows a word to be automatically targeted in the text and replaced with its
156 synonym. In practice, this dictionary is derived from the valence dictionary used in Luc et al. (2020).
157 Moreover, with Part-Of-Speech tagging using the IRaMuTeQ© software, the detection of all the
158 synonyms that were not already taken into account in the dictionary is easier.

159 Finally, we advocate attaching the terms "too much", "JAR" and "not enough" to the attributes
160 they refer to, in order to achieve a straightforward assessment of the assets and weaknesses of the
161 products. This means that, in each comment, the attribute and its associated quantifier form a single
162 word (e.g., "too_much_creamy", "just_sweet_enough").

163

164 2.3. Analysis at the comment level: valency scores

165

166 The analysis of the data at a comment level (block 3 in Figure 1) consists of assessing the positive
167 or negative aspect of each comment by means of a valency score. The use of particular terms in the
168 Free JAR comments and, in particular, the terms "too much", "not enough" and "just about right",
169 highlights the products' assets and weaknesses. Thus, the respondents convey a valence, be it
170 positive or negative, depending on which words are used. A strategy of assigning scores to comments
171 was discussed in Luc et al. (2020). However, since this strategy was based on a lexical, dictionary-

172 based, approach, it was very time-consuming. To cope with this impediment, we propose to compute
173 valency scores from the outputs of a classifier applied to the comments. This strategy of analysis
174 draws from procedures akin to natural language processing (NLP; Manning & Schütze, 1999; Liddy,
175 2001). NLP is a subfield of linguistics, computer sciences and artificial intelligence that makes it
176 possible to analyze large amounts of human language data.

177 Central to our approach is the investigation of the relationships between the free comments and
178 their associated hedonic category. More precisely, we advocate using a supervised classification
179 approach whose aim is to find rules for assigning objects to predefined categories (Carrizosa &
180 Romero Morales, 2013). This strategy is based on setting up a classifier which, for Free JAR data, is
181 trained to assign each comment to one of the three hedonic categories. Examples of classifiers are
182 Random Forest (RF) (Breiman, 1984, 2001) or Support Vector Machine (SVM) (Cortes & Vapnik,
183 1995). In this paper, the RF classifier is applied since it is commonly used in the literature to analyze
184 textual data (Amato et al., 2021; De Clercq et al., 2019).

185 RF is an ensemble learning method made of a large number, say 500, of decision trees. These
186 trees are created according to a particularly efficient strategy aimed at increasing the diversity
187 between them, firstly by fitting each tree on a bootstrap replicate (Efron & Tibshirani, 1993) of the
188 full set of samples and secondly, by selecting a small random subset of words from the total words
189 and using only this subset to search for the best split. For the classification task considered in our
190 case study, the output of RF is the membership probabilities of a comment to each of the three
191 hedonic categories, or alternatively the category selected by the majority of the trees in the forest
192 for the considered comment. The use of RF classifier involves several parameters that need to be
193 tuned in order to achieve the best fit. Such parameters are the number of trees, the number of
194 words randomly selected at each split of each tree (so-called *mtry* parameter) and the number of
195 observations in the terminal nodes. The choice of these parameters can be achieved by assessing the
196 classifier performance on the basis of the out-of-bag observations, which are the observations set
197 aside when the bootstrap sample associated to each tree is selected (for more details see for
198 instance Breiman (2001), Granitto et al. (2007), Vigneau et al. (2018)).

199
200 Thus, by applying the RF classifier to a given Free JAR comment, we obtain membership
201 probabilities to each of the three hedonic categories. We define the valency score of a given
202 comment as the difference between the membership probability that the comment belongs to the
203 category "I like very much" and the membership probability that the comment belongs to the
204 category "I don't like". Therefore, this score ranges between -1 and 1 and the higher it is, the more
205 positive is the comment.

206 In order to check the relevance of this procedure, we consider two indices: (i) The accuracy
207 index, which is defined as the ratio of the number of correct comment classifications to the total
208 number of comments. This index is a measure of the performance of the classifier applied to the Free
209 JAR data, and can be also expressed for each category separately; (ii) The correlation ratio (η)
210 (Pearson, 1911) which measures the intensity of the link between the valency scores and the hedonic
211 categories. A correlation ratio close to 1 means that the valency scores within each hedonic category
212 are very similar whereas the valency scores from different categories are different. Formally, it is
213 equal to the square root of the ratio of the between categories sum of squares to the total sum of
214 squares of the valency scores.

215 The valency scores can serve different purposes. For example, an internal preference mapping
216 (Carroll, 1972; Danzart, 2009; Greenhoff & MacFie, 1994) based on these scores could be carried out.
217 This consists in performing an unstandardized Principal Component Analysis (PCA) on the table
218 whose rows are the products and whose columns are the respondents. Each entry of this table is the
219 valency score associated with the description of a given respondent and product.

220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264

2.4. Analysis at the word level: interpretability and the LIME algorithm

Once the positive or negative nature of each comment is assessed by means of a valency score, we seek to unveil the reasons why a given comment is positive or negative. This is achieved by highlighting those words in the comment that entail a positive valence and those words that entail a negative valence. To address this goal, we use the concept of interpretability to pinpoint the words that have a significant impact on the classification process (block 4 in Figure 1).

In particular, the LIME algorithm (for Local Interpretable Model-agnostic Explanations) is used. It is defined as "an algorithm that can explain the predictions of any classifier or regressor in a faithful way, by approximating it locally with an interpretable model" (Ribeiro et al., 2016). Thus, although the interpretation of the original model such as RF may not be intuitive, LIME sets up a model that is locally consistent with the classifier, and easier to interpret.

The LIME algorithm involves five steps (Molnar, 2019). In our context, these steps can be described as follows:

1. Select the comment of interest, for which an explanation regarding the prediction of the classifier is sought.
2. Disrupt the comment. This means that new comments are created by switching some words from present to absent or vice versa (see Table 3 for an illustration). The classification predictions for these new disrupted comments are computed using the RF classifier, trained on the comments and their associated hedonic categories from all the respondents.
3. Compute a proximity score for each disrupted comment as the percentage of common words between this disrupted comment and the original comment.
4. Train an easy-to-interpret model (e.g., linear model) on these new data, i.e. the disrupted comments associated with their hedonic category predicted in step 2, weighted by the proximity scores. This model seeks to predict the membership probability of each disrupted comment to the category "I like very much" from the binary data that reflect the absence or the presence (0/1) of the different words that compose the original comment (see Table 3).
5. The local model yields regression coefficients associated with the various words.

In practice, several hyperparameters should be adjusted before running the LIME algorithm. For example, the "label" argument allows us choosing the category for which an explanation is sought. Obviously, in our case, we chose the "I like very much" category in order to highlight the drivers of liking and disliking. LIME algorithm also offers a range of easy-to-interpret models to be selected from the scikit-learn Python library (https://scikit-learn.org/stable/modules/linear_model.html). The selected linear model should be indicated through the "model" argument, and will be used in step 4 of the LIME procedure. In the present paper, we selected Ridge regression (Hoerl & Kennard, 1970a, 1970b) as it is one of the simplest models among those handled by the LIME function.

As an illustration, we give in Table 3 the predicted class regarding a specific comment together with the predicted classes of disrupted comments.

<i>too_creamy</i>	<i>just_sweet_enough</i>	<i>too_soft</i>	Predicted class	Probability "I like very much"	Proximity score
-------------------	--------------------------	-----------------	-----------------	--------------------------------	-----------------

1	1	1
1	1	1
1	1	0
1	0	1
0	1	0

3	0.22	0.265
3	0.22	0.266
2	0.45	0.267
3	0.13	0.268
1	0.87	0.269
		0.270
		0.271
		0.272

A
s
state
d
abov
e,
LIME
algo

273 rithm yields coefficients associated with the various words that reflect their contribution to the
274 assignment of each comment to the hedonic category "I like very much". In the following, these

Table 3: Example of disruptions for the comment "too_creamy, just_sweet_enough, too_soft": "1" means that the word is present and "0" means that the word is absent. The right part of the table gives the predicted classes and the membership probabilities to the category "I like very much" resulting from RF classifier. The last column gives the proximity scores assigned to each disrupted comment that reflect the extent to which each disrupted comment is close to the original comment.

275 coefficients will be referred to as the "word weights".

276 It is noteworthy that a given word can be used by different respondents to describe different
277 products. This entails that this word will have as many weights as the number of times it has been
278 used in the various comments.

279 For each product, the weights associated with each word that has been used to describe it are
280 averaged over all the respondents. The number of times the word has been used to describe the
281 product under study is also recorded. These two quantities will be depicted in a graphical display
282 with the x-axis representing the mean weight of the various words and the y-axis, their numbers of
283 occurrences. In practice, it seems reasonable to focus only on the most frequently cited words. We
284 chose to represent here only the words cited by more than 5% of the respondents. Thus, the liking
285 and disliking drivers can be highlighted by focusing on the words situated in the upper right side of
286 the graph for the liking drivers, and in the upper left side of the graph for the disliking drivers.

287 2.5. Case study

288 The data collection took place in downtown Bordeaux, France. Eight French pressed cheeses
289 were assessed, including one cheese (Cantal) which was replicated (Table 4). These cheeses are
290 among the most commonly available and commonly consumed pressed cheeses in France. All the
291 cheeses are made in France, under a protected designation of origin (PDO) label (except the
292 Emmental) and were purchased at the same supermarket.

Cheese	Abbreviation	Type
Beaufort	B	Cooked pressed cheese
Cantal	C	Uncooked pressed cheese
Cantal (Replic.)	CR	Uncooked pressed cheese

Comté	CE	Cooked pressed cheese
Emmental	E	Cooked pressed cheese
Morbier	M	Uncooked pressed cheese
Reblochon	R	Uncooked pressed cheese
Saint-Nectaire	S	Uncooked pressed cheese

Table 1: The eight French cheeses used in the case study and their abbreviation. The type of pressed cheese is also indicated.

A panel of 77 respondents took part in the experiment. They were recruited in the street by a team of investigators and claimed to be regular consumers of pressed cheeses. Among these respondents, 39% were male and 61% were female. They were aged between 18 and 73 years.

Clearly the recruitment was a convenience sample since the experiment was set up with the aim of demonstrating the interest of Free JAR procedure. Therefore, the focus is put on the statistical analysis.

The data collection took place in a test room. Upon arrival in the room, each respondent was seated in an individual booth and a cheese tray was brought to them. The eight products were presented in the form of two cubes of dimensions 1 x 1 x 1.5 cm. They were arranged on the tray according to a balanced experimental design in order to avoid carry-over effect. The instruction given to the respondents was to use the first cheese cube of each product for the sorting task, and the second one for the description task. Each respondent has neutral crackers and Evian water in order to neutralize their palate after tasting each cheese sample. Data were collected in handwritten form on paper sheets.

The statistical analyses were performed within the R software, version 4.0.3 (R Core Team, 2020) and the Python software version 3.7.4 (Van Rossum & Drake Jr, 1995). In particular, the *FactoMineR* R package (Lê et al., 2008) version 2.4, the *randomForest* R package (Liaw & Wiener, 2002) version 4.6, the *e1071* R package (Meyer et al., 1999) version 1.7 and the *lime* Python package (Ribeiro et al., 2016) version 0.2.0.1 were used.

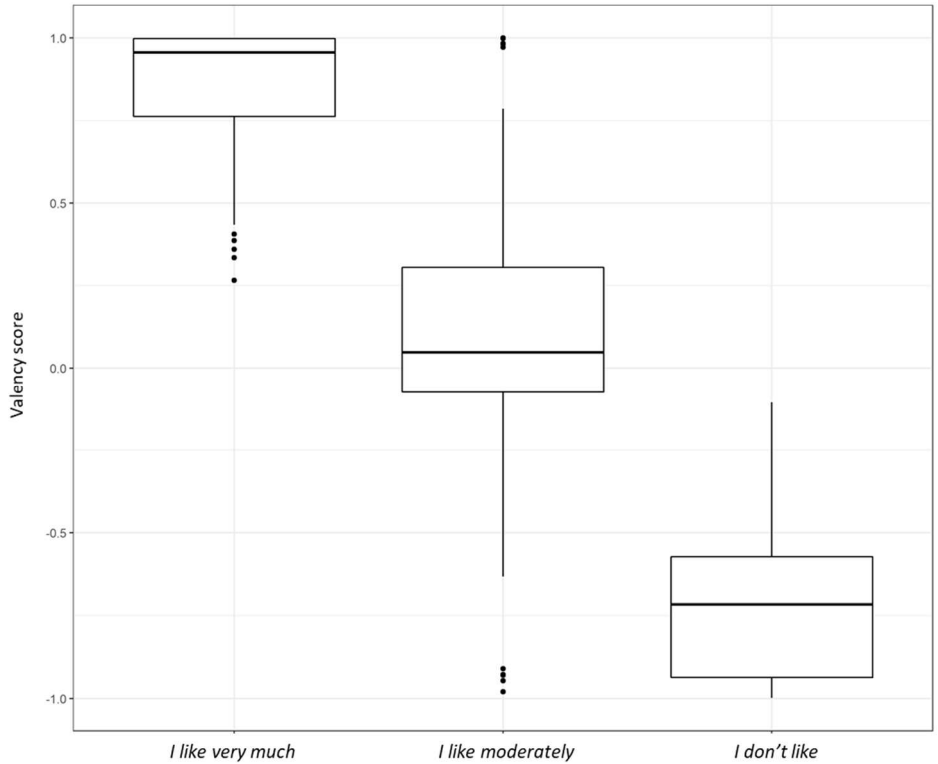
3. Results

The data collection yielded 616 Free JAR comments; each of them being associated with a hedonic category. After preprocessing the data following the procedure described in section 2.2., a document-term matrix was created, composed of 616 rows that correspond to the Free JAR comments and 400 columns that correspond to the attached words (e.g., "too_much_creamy", "just_sweet_enough") used by the respondents to describe the products.

3.1. Analysis at the comment level

The valency scores are calculated for each comment following the procedure described in section 2.3. In practice, the tuning of the RF model led to a number of trees of 500, a number of words tested

353 at each split (*mtry* parameter) of 150 and a number of observations in the terminal nodes of 1. The
354 default value (20) of the *mtry* parameter did not provide the best performance.



375
376 *Figure 2: Boxplots representing the valency scores, obtained by means of RF classifier, according*
377 *to the hedonic categories*

379 Boxplots that represent the distribution of the valency scores according to the hedonic
380 categories are plotted in Figure 2. Although these boxplots show an overlap of the valency scores for
381 the three hedonic categories, it is clear that the valency scores for the comments in the category "I
382 like very much" tend to be larger than those of the "I like moderately" category. Similarly, the valency
383 scores of the comments from this latter category tend to be larger than those of the "I don't like"
384 category. A one-way ANOVA indicated that there is a significant effect of the hedonic categories on
385 the valency scores. The correlation ratio (*i.e.*, square root of the between to total sum of squares
386 ratio) is equal to 0.92. Furthermore, the classification accuracy (*i.e.*, percentage of correct
387 classification) is equal to 0.93, globally. For each hedonic category, namely "I like very much", "I like
388 moderately and "I don't like", the classification accuracy is equal to 0.98, 0.90 and 0.89, respectively.

390 As stated above, the valency scores can be used to set up a preference mapping of the products.
391 For this effect, we considered the data table whose rows are the products and whose columns are
392 the various respondents. The entries of this data table are the valency scores associated with the
393 comments provided by each respondent to describe each product. The resulting table was submitted
394 to a non-standardized Principal Components Analysis. Figure 3 shows the simultaneous
395 representations of the products and the respondents.

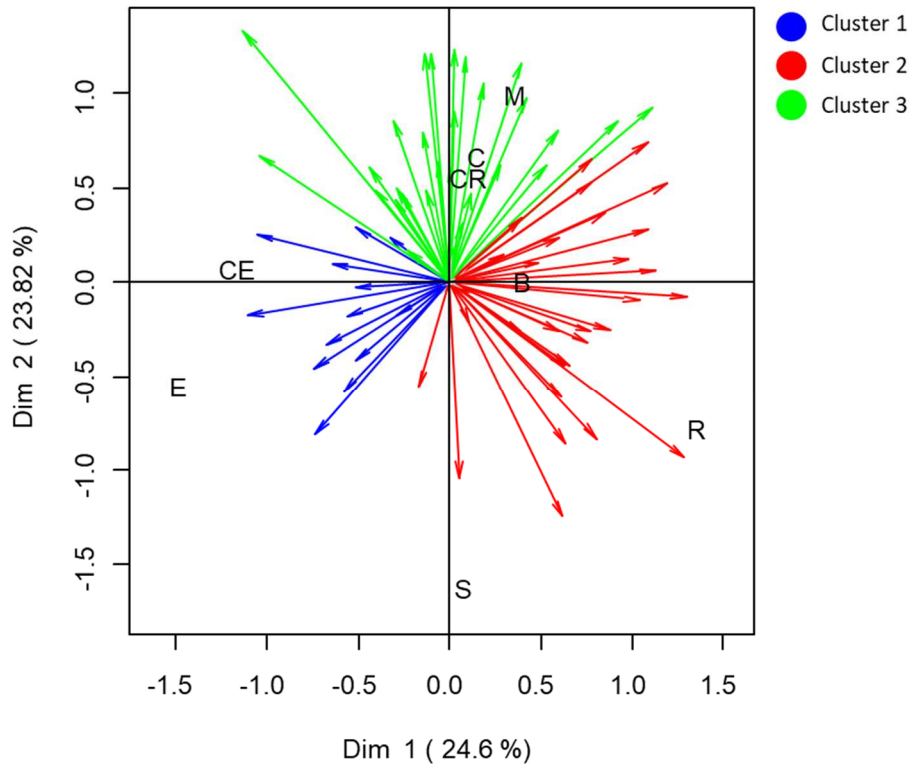


Figure 3: Biplot resulting from a PCA on the valency scores from the Free JAR comments. Clusters of respondents highlighted by means of a CLV procedure are indicated.

A segmentation of the respondents by means of the Clustering of variables around Latent Components, CLV, method (Vigneau et al., 2015; Vigneau & Qannari, 2003) was also performed, and the different clusters of respondents are highlighted in the biplot (Figure 3).

Regarding the product configuration, we can notice that, on the one hand, both the Cantal (C and CR) are close to each other and close to the Beaufort (B) and the Morbier (M). On the other hand, the Emmental (E) and the Comté (CE) seem to be close, as well as the Saint-Nectaire (S) and the Reblochon (R).

The segmentation of the respondents reveals three clusters composed of 15, 30 and 32 respondents, respectively. The respondents that belong to each of them seem to have described a particular set of products in a similar way. The respondents in the first cluster have their arrows directed towards the Emmental (E) and the Comté (CE) cheeses. This reflects the fact that they like these cheeses and, by opposition, dislike the Reblochon (R). The respondents in the second cluster tend to appreciate more the Saint-Nectaire (S), the Reblochon (R) and the Beaufort (B) cheeses. The respondents in the third cluster appreciate more the Cantal (C and CR) and the Morbier (M) cheeses.

3.2. Analysis at the word level

The LIME algorithm is applied to each Free JAR comment in order to assign a weight to each of the words present in the Free JAR comment. For each cluster of respondents highlighted in section 3.1. and for each product, the mean weight of each word and its number of occurrences were computed. As mentioned above, these quantities are depicted by means of a graphical display showing the mean weights of the words versus their number of occurrences. We have chosen to represent only the words cited by more than 5% of the panel size, which corresponds to the words whose number of occurrences is larger than 4. For the sake of saving space, we show herein the

449 results for products C, CR and S, namely the Cantal cheese, its replicate and the Saint-Nectaire
 450 cheese. Drivers of liking and disliking for each of these three products resulting from this analysis are
 451 displayed in Figure 4, according to the clusters of respondents.

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

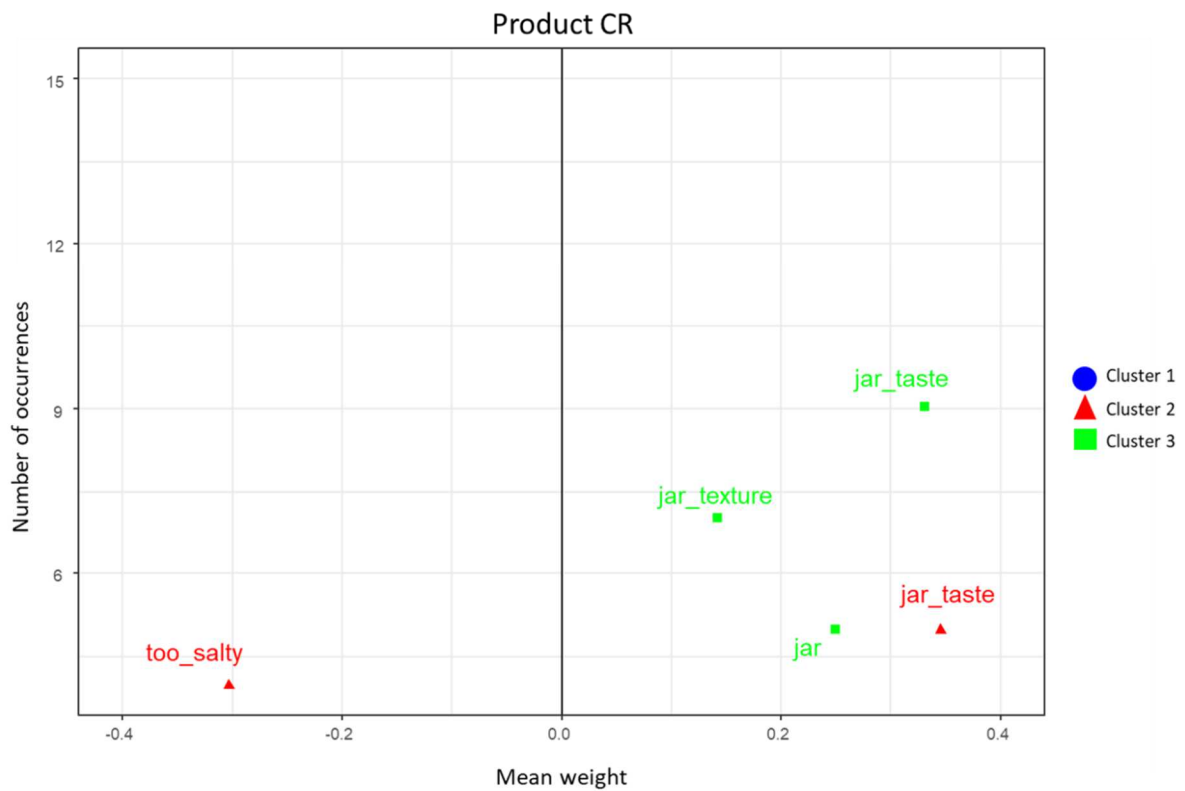
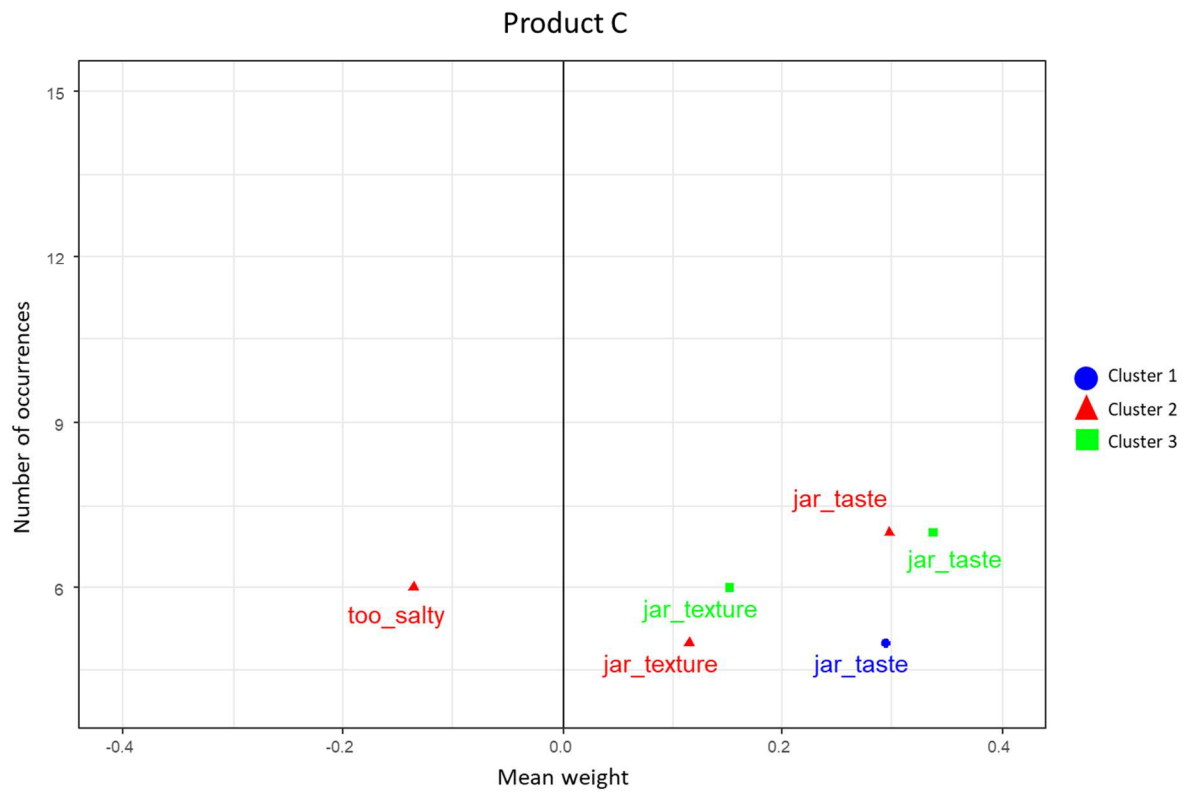
478

479

480

481

482



483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522

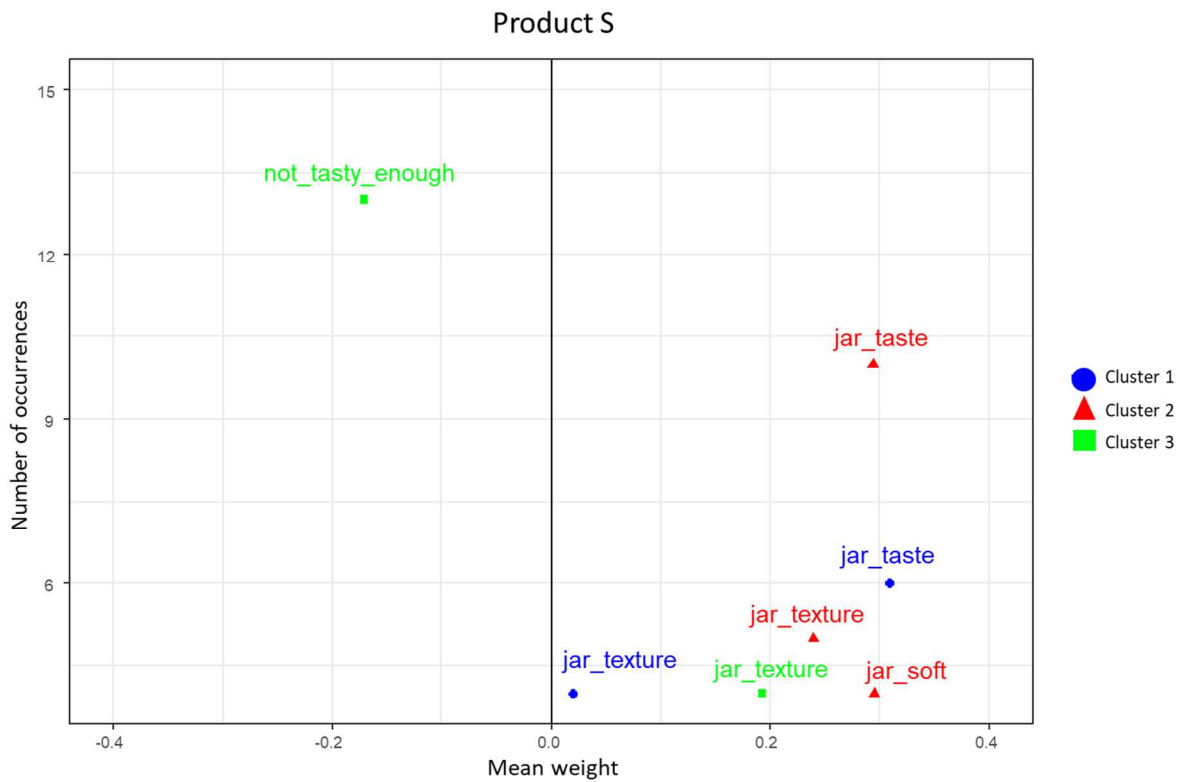


Figure 4: Drivers of liking (weight > 0) and disliking (weight < 0) highlighted for the Cantal (C and CR) and the Saint-Nectaire (S) cheeses, through an interpretability procedure based on the LIME algorithm. Only words cited more than 4 times (more than 5% of the panel size) are shown.

It is noteworthy that the sparseness among respondents' assessments entails that different findings can be drawn depending on which cluster is considered. In particular, the analysis of the comments relative to the Cantal (C and CR) cheeses shows that respondents in cluster 2 are the only ones that highlight disliking drivers for these products. In particular, the Cantal is perceived as too salty for them. Overall, comparing the results for product C and CR in Figure 4 reveals that the respondents appear consistent in their assessments, despite their small number. These findings are also consistent with the biplot shown in Figure 3 as the Cantal (C and CR) cheeses are close to each other and opposed to other products that are preferred by respondents in cluster 2.

Similarly, the analysis of the comments related to the Saint-Nectaire (S) cheese mainly reveals liking drivers from respondents in cluster 1 and 2. It also reveals a clear disliking driver from respondents in cluster 3. Considering the mean weights of the drivers and their number of occurrences, we can state that the Saint-Nectaire (S) cheese appears to be just right in terms of texture for the respondents in the three clusters, but tends to be perceived as not tasty enough for the respondents in cluster 3. This finding is consistent with the biplot shown in Figure 3 as the Saint-Nectaire (S) is opposed to products preferred by respondents in cluster 3.

4. Discussion and concluding remarks

The findings from the approach developed in this paper are in line with the inherent properties of the products and with the findings from a previous study on the same set of products (Luc et al., 2021). By comparison to this latter study, the approach adopted herein is much less time consuming since it relies on easier treatments of the textual data. The machine learning approach applied to Free JAR data allows us to take advantage of the link between the hedonic categorization and the textual description of the products, in order to interpret the comment of interest. By implementing

523 an RF classifier, a valency score was obtained for each comment, reflecting its positive or negative
524 overtone. These quantitative scores allow us to broaden the scope of the analyses to be performed
525 on the textual data. An example of such analyses is the internal preference mapping, as illustrated in
526 the present paper. It is also noteworthy that this approach can be performed on data from any
527 language, since the classifier is only based on the link between word presence or absence and
528 hedonic categories. This is another clear advantage of using a machine learning approach over a
529 lexical approach as developed in Luc et al. (2020).

530 Of paramount interest is the concept of interpretability and the LIME algorithm. By applying this
531 algorithm to Free JAR data and taking into account the heterogeneity of the panel of respondents,
532 we were able to highlight drivers of liking and disliking for the various products. It took us around 6
533 minutes to run this algorithm and extract all the word weights on our whole dataset composed of
534 616 comments. The results regarding the eight products are available as supplementary material,
535 and their analysis allowed us to state that, for instance, both the Cantal (C and CR) cheeses are
536 perceived as too salty for some respondents, while the Beaufort (B) and the Morbier (M) cheeses are
537 perceived as too strong for others. The Reblochon (R), the Saint-Nectaire (S), and the Emmental (E)
538 cheeses are perceived as not tasty enough for some respondents, while the Comté (CE) cheese is
539 perceived as both too bland for some respondents, and just right in terms of taste for others. Thus,
540 LIME appears to be a reliable way to obtain drivers of liking and disliking from labeled textual data.

541 It is clear that the number of occurrences of each word is an important parameter. However,
542 since the respondents use their own words spontaneously, the number of unique words or words
543 with relatively small frequency is likely to be large. In the present case study, we chose not to
544 perform heavy data pre-processing such as grouping attributes according to their meanings, and we
545 made the choice to represent only the words that have been used more than four times.

546 The choice of RF as a classifier was grounded on the fact that it is very popular in text mining.
547 However, by way of comparing results, we also applied an SVM classifier. It turned out that RF
548 outperformed SVM as reflected by the performance indices, namely correlation ratio (0.92 vs 0.68)
549 and accuracy (0.93 vs 0.56).

550 It is noteworthy that the classifier performance depends on the textual preprocessing step. In
551 fact, less involving preprocessing was also tested by simply removing the primers and stopwords. It
552 turned out that the performance in terms of correlation ratio and accuracy was less satisfactory. This
553 demonstrates the importance of the combination of data preprocessing and machine learning
554 algorithms to improve the performance of the overall approach. This is particularly the case in the
555 field of sensory analysis where the amount of data collected is not generally large enough to perform
556 robust analyses based on machine learning techniques.

557
558 As a general conclusion, this study shows the range of possibilities offered by combining hedonic
559 categorization and textual description in a free comment procedure, such as in the Free JAR
560 procedure. Whereas some steps of the procedure developed in this paper are clearly specific to the
561 analysis of Free JAR data (e.g., the synonymous grouping of JAR terms and the highlighting of drivers
562 of liking and disliking), we believe that the proposed procedure could be applied to any labeled
563 textual data.

564 Further studies should be undertaken to confirm the value of this strategy of analysis on labeled
565 data pertaining to other products, and to better understand its limitations and find ways to cope with
566 them. Automating the various tasks, especially during the preprocessing phase, is of paramount
567 interest both to save time and cost and to improve the model performance. Currently, major
568 advances are being made in machine learning and text mining, and this will certainly enable us to
569 further improve the general strategy of analysis proposed herein.

570 **Acknowledgments**

571

572 The authors thank the two anonymous reviewers for their insightful comments that allowed us to
573 significantly improve the quality of the present paper.

574 **References**

- 575
- 576 Amato, F., Coppolino, L., Cozzolino, G., Mazzeo, G., Moscato, F., & Nardone, R. (2021). Enhancing
577 random forest classification with NLP in DAMEH: A system for DATA Management in eHealth
578 Domain. *Neurocomputing*, 444, 79–91. <https://doi.org/10.1016/j.neucom.2020.08.091>
- 579 Ares, G., Giménez, A., Barreiro, C., & Gámbaro, A. (2010). Use of an open-ended question to identify
580 drivers of liking of milk desserts. Comparison with preference mapping techniques. *Food*
581 *Quality and Preference*, 21(3), 286–294. <https://doi.org/10.1016/j.foodqual.2009.05.006>
- 582 Ares, G., & Jaeger, S. R. (2015). Check-all-that-apply (CATA) questions with consumers in practice:
583 Experimental considerations and impact on outcome. In *Rapid Sensory Profiling Techniques*
584 (pp. 227–245). Elsevier. <https://doi.org/10.1533/9781782422587.2.227>
- 585 Bécue-Bertaut, M., Álvarez-Esteban, R., & Pagès, J. (2008). Rating of products through scores and
586 free-text assertions: Comparing and combining both. *Food Quality and Preference*, 19(1),
587 122–134. <https://doi.org/10.1016/j.foodqual.2007.07.006>
- 588 Borko, H. (1965). A Factor Analytically Derived Classification System for Psychological Reports.
589 *Perceptual and Motor Skills*, 20(2), 393–406. <https://doi.org/10.2466/pms.1965.20.2.393>
- 590 Breiman, L. (Ed.). (1984). *Classification and regression trees* (Repr). Chapman & Hall [u.a.].
- 591 Breiman, L. (2001). Random Forest. *Machine Learning*, 45(1), 5–32.
592 <https://doi.org/10.1023/A:1010933404324>
- 593 Carrizosa, E., & Romero Morales, D. (2013). Supervised classification and mathematical optimization.
594 *Computers & Operations Research*, 40(1), 150–165.
595 <https://doi.org/10.1016/j.cor.2012.05.015>
- 596 Carroll, J. D. (1972). Individual differences and multidimensional scaling. *R.N. Shepard, A.K. Romney,*
597 *and S.B. Nerlove (Eds) Multidimensional Scaling: Theory and Applications in the Behavioral*
598 *Sciences*, 1, 105–155.
- 599 Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
600 <https://doi.org/10.1007/BF00994018>

601 Danzart, M. (2009). Cartographie des préférences. In *Evaluation sensorielle—Manuel*
602 *méthodologique* (3rd ed., pp. 443–449). Lavoisier, SSHA.

603 De Clercq, D., Diop, N.-F., Jain, D., Tan, B., & Wen, Z. (2019). Multi-label classification and interactive
604 NLP-based visualization of electric vehicle patent data. *World Patent Information*, *58*,
605 101903. <https://doi.org/10.1016/j.wpi.2019.101903>

606 Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Springer US.
607 <https://doi.org/10.1007/978-1-4899-4541-9>

608 Granitto, P. M., Gasperi, F., Biasioli, F., Trainotti, E., & Furlanello, C. (2007). Modern data mining tools
609 in descriptive sensory analysis: A case study with a Random forest approach. *Food Quality*
610 *and Preference*, *18*(4), 681–689. <https://doi.org/10.1016/j.foodqual.2006.11.001>

611 Greenhoff, K., & MacFie, H. J. H. (1994). Preference mapping in practice. In H. J. H. MacFie & D. M. H.
612 Thomson (Eds.), *Measurement of Food Preferences* (pp. 137–166). Springer US.
613 https://doi.org/10.1007/978-1-4615-2171-6_6

614 Gutjar, S., de Graaf, C., Kooijman, V., de Wijk, R. A., Nys, A., ter Horst, G. J., & Jager, G. (2015). The
615 role of emotions in food choice and liking. *Food Research International*, *76*, 216–223.
616 <https://doi.org/10.1016/j.foodres.2014.12.022>

617 Hoerl, A. E., & Kennard, R. W. (1970a). Ridge Regression: Applications to Nonorthogonal Problems.
618 *Technometrics*, *12*(1), 69–82. <https://doi.org/10.1080/00401706.1970.10488635>

619 Hoerl, A. E., & Kennard, R. W. (1970b). Ridge Regression: Biased Estimation for Nonorthogonal
620 Problems. *Technometrics*, *12*(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>

621 King, S. C., Meiselman, H. L., & Thomas Carr, B. (2013). Measuring emotions associated with foods:
622 Important elements of questionnaire and test design. *Food Quality and Preference*, *28*(1), 8–
623 16. <https://doi.org/10.1016/j.foodqual.2012.08.007>

624 Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of*
625 *Statistical Software*. *Journal of Statistical Software*, 1–18.

626 Lebart, L., Salem, A., & Berry, L. (1998). *Exploring Textual Data* (Vol. 4). Springer Netherlands.
627 <https://doi.org/10.1007/978-94-017-1525-6>

628 Lewin, K. (1935). *A dynamic theory of personality*. McGraw-Hill Book Company, Inc.

629 Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*.
630 <https://CRAN.R-project.org/doc/Rnews/>

631 Liddy, E. D. (2001). Natural Language Processing. In *Encyclopedia of Library and Information Science*
632 (2nd ed., p. 15). Marcel Decker, Inc.

633 Luc, A., Lê, S., & Philippe, M. (2020). Nudging consumers for relevant data using Free JAR profiling: An
634 application to product development. *Food Quality and Preference*, 79, 103751.
635 <https://doi.org/10.1016/j.foodqual.2019.103751>

636 Luc, A., Le, S., Philippe, M., Qannari, E. M., & Vigneau, E. (2021). Free JAR experiment: Data analysis
637 and comparison with JAR task. *Food Quality and Preference*, 104453.
638 <https://doi.org/10.1016/j.foodqual.2021.104453>

639 Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT
640 Press.

641 Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., & Lin, C. (1999). *E1071:*
642 *Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071),*
643 *TU Wien*.

644 Miller, T. (2018). Explanation in Artificial Intelligence: Insights from the Social Sciences.
645 *ArXiv:1706.07269 [Cs]*. <http://arxiv.org/abs/1706.07269>

646 Mohammad, S. M. (2016). Sentiment Analysis. In *Emotion Measurement* (pp. 201–237). Elsevier.
647 <https://doi.org/10.1016/B978-0-08-100508-8.00009-6>

648 Molnar, C. (2019). *Interpretable Machine Learning A Guide for Making Black Box Models Explainable*.

649 Pearson, K. (1911). On a correction to be made to the Correlation Ratio. *Biometrika*, 8(1/2), 254–256.
650 <https://doi.org/10.2307/2331454>

651 Piqueras-Fizman, B. (2015). Open-ended questions in sensory testing practice. In *Rapid Sensory*
652 *Profiling Techniques* (pp. 247–267). Elsevier. <https://doi.org/10.1533/9781782422587.2.247>

653 R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for
654 Statistical Computing. <https://www.R-project.org/>

655 Ratinaud, P. (2014). *IRaMuTeQ : Interface de R pour les analyses Multidimensionnelles de Textes et de*
656 *Questionnaires* (0.7 alpha 2) [Computer software].

657 Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). ‘Why Should I Trust You?’: Explaining the Predictions
658 of Any Classifier. *ArXiv:1602.04938 [Cs, Stat]*. <http://arxiv.org/abs/1602.04938>

659 Rothman, L., & Parker, M. J. (Eds.). (2009). Structure and Use of Just-About-Right Scales. In *Just-*
660 *About-Right (JAR) Scales: Design, Usage, Benefits, and Risks* (pp. 1–13). ASTM International.
661 <https://doi.org/10.1520/MNL11481M>

662 ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey responses: A
663 complementary method to preference mapping. *Food Quality and Preference, 14*(1), 43–52.
664 [https://doi.org/10.1016/S0950-3293\(02\)00011-3](https://doi.org/10.1016/S0950-3293(02)00011-3)

665 Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial*.

666 Vigneau, E., Chen, M., & Qannari, E., Mostafa. (2015). ClustVarLV: An R Package for the Clustering of
667 Variables Around Latent Variables. *The R Journal, 7*(2), 134–148.
668 <https://doi.org/10.32614/RJ-2015-026>

669 Vigneau, E., Courcoux, P., Symoneaux, R., Guérin, L., & Villière, A. (2018). Random forests: A machine
670 learning methodology to highlight the volatile organic compounds involved in olfactory
671 perception. *Food Quality and Preference, 68*, 135–145.
672 <https://doi.org/10.1016/j.foodqual.2018.02.008>

673 Vigneau, E., & Qannari, E. M. (2003). Clustering of Variables Around Latent Components.
674 *Communications in Statistics - Simulation and Computation, 32*(4), 1131–1150.
675 <https://doi.org/10.1081/SAC-120023882>

676 Visalli, M., Mahieu, B., Thomas, A., & Schlich, P. (2020). Automated sentiment analysis of Free-
677 Comment: An indirect liking measurement? *Food Quality and Preference*, 82, 103888.
678 <https://doi.org/10.1016/j.foodqual.2020.103888>
679