Blending sentiment and topic analysis to explain discrete emotions from online fitness center customer reviews

Combinación del análisis del sentimiento y del contenido para explicar emociones discretas a partir de reseñas online de clientes de gimnasios

Authors

Enrique Alcántara Alcover [1]
Antonio Hernandez-Martin [2]
Ricardo Cuevas [1]
Daniel Duclos-Bastías [3, 4]

[1] Universidad de Castilla-La Mancha (Spain)
[2] Universidad Europea de Madrid (Spain)
[3] Pontificia Universidad Católica de Valparaíso, Valparaíso (Chile)
[4] Universidad Alfonso X el Sabio (UAX) (Spain)

Corresponding author:
Daniel Duclos-Bastías
daniel.duclos@pucv.cl

Received: 27-01-26
Accepted: 04-03-26

How to cite in APA

Alcántara Alcover, E., Hernandez-Martin, A., Cuevas, R., & Duclos-Bastías, D. (2026). Blending sentiment and topic analysis to explain discrete emotions from online fitness center customer reviews. *Retos, 78*, 767-783. https://doi.org/10.47197/retos.v78.118658

Abstract

Introduction: Emotions elicited by services influence customer satisfaction and overall experience. Online reviews provide a valuable source for identifying these emotional patterns through text-analysis techniques.
Objective: To examine how topic relevance and sentiment expressed in digital reviews predict discrete emotions among gym users, proposing an approach based on "polarized topics."
Methodology: A total of 3,250 reviews from 38 Spanish gyms were analyzed. The study employed a mixed-methods approach combining topic modeling using LDA, sentiment analysis with TextBlob, and machine-learning algorithms (XGBoost), integrated with SHAP explainability techniques. The interaction between topics and sentiment polarity was used to predict ten discrete emotions, including joy, anger, sadness, and trust.
Results: Staff friendliness, value for money, and hygiene emerged as highly predictive topics. Positive evaluations of staff increased emotions such as joy and trust, whereas comments related to COVID-related absences were associated with higher levels of anger and sadness. The "polarized topics" approach yielded strong emotional classification performance, achieving F1-scores above 0.84 for most emotions.
Discussion: The findings show that the combination of topic modeling, sentiment analysis, and explainable AI enables precise identification of which service attributes trigger specific emotions, offering a useful interpretive framework for managers in the fitness sector.
Conclusions: The proposed method constitutes a scalable and transparent approach for predicting discrete emotions in reviews. Its application can support improvements in service design and emotional alignment in organizations oriented toward wellbeing.

Keywords

Emotions; fitness; machine learning; online reviews; stars-ranking.

Resumen

Introducción: Las emociones generadas por los servicios influyen en la satisfacción y la experiencia del cliente. Las reseñas online permiten identificar estos patrones emocionales mediante técnicas de análisis de texto.
Objetivo: Examinar cómo la relevancia temática y el sentimiento presentes en opiniones digitales predicen emociones discretas de usuarios de gimnasios, proponiendo un enfoque basado en "temas polarizados".
Metodología: Se recopilaron 3.250 reseñas de 38 gimnasios españoles. El estudio aplicó un método mixto que combinó modelado temático mediante LDA, análisis de sentimiento con TextBlob y algoritmos de aprendizaje automático (XGBoost), integrados con técnicas de explicabilidad SHAP. La interacción entre temas y polaridad del sentimiento se empleó para predecir diez emociones discretas, como alegría, ira, tristeza o confianza.
Resultados: La amabilidad del personal, la relación calidad-precio y la higiene emergieron como temas especialmente predictivos. Las valoraciones positivas relacionadas con el personal incrementaron emociones como alegría y confianza, mientras que comentarios sobre bajas por COVID se asociaron con mayores niveles de ira y tristeza. El enfoque de "temas polarizados" generó una clasificación emocional sólida, alcanzando puntuaciones F1 superiores a 0,84 en la mayoría de las emociones.
Discusión: Los hallazgos muestran que la combinación de modelado temático, sentimiento y técnicas explicables permite identificar con precisión qué aspectos del servicio desencadenan emociones concretas, proporcionando un marco interpretativo útil para gestores del sector fitness.
Conclusiones: El método propuesto constituye una aproximación escalable y transparente para predecir emociones discretas en reseñas. Su aplicación puede mejorar el diseño del servicio y la alineación emocional en organizaciones orientadas al bienestar.

Palabras clave

Aprendizaje automático; emociones; fitness; reseñas online; valoraciones con estrellas.

Introduction

Extensive literature shows that product and service characteristics are relevant in triggering customers emotions that influence their behavior and consumption experience (Ogruk, Anderson & Nacass, 2018; Kranzbühler et al., 2020; Guo et al., 2014). Among these research's, it has been found, for example, that the appropriate combination of environment elements triggers the positive emotional state of a customer that leads to a pleasant, memorable experience, more time and money spent in a store (Ogruk, Anderson & Nacass, 2018). The relationship between physiological measurements of emotional experiences (eye tracking, questionnaires, pulse, galvanic skin response, etc.) and customer focus on design elements has also been demonstrated in experimental designs during shopping process on e-commerce website pages (Guo et al., 2014).

Emotional Design (ED) is a discipline that aims at establishing the relationship between the characteristics of products and services and the emotions they elicit in customers (Jiao & Qu, 2019; Chiu & Lin, 2018; Li, Go & Jin, 2020; Lui, Cui & He, 2021). Currently, the analysis of the emotional content of user online reviews is gaining traction as an effective alternative in terms of cost, time and data scale for research in many realms or to speed up the process of product and service development (Chaklader & Parkison, 2017; Chiu & Lin, 2018; Jia, 2019). There is a growing number of publications analyzing how emotions readers perceive the content of online reviews influence on their purchase decisions and attitudes (Yi et al., 2023; Kim & Gupta, 2012; Guo, Wang & Wu, 2020; Mao et al., 2019; Fan et al., 2022; Lelieveld & Hendriks 2021; Fernandes et al., 2022). However, its application in ED is still limited and research in this field usually relies on traditional methods such as interviews or surveys (Jiao & Qu, 2019; Chiu & Lin, 2018; Li, Go & Jin, 2020; Lui, Cui & He, 2021). In this sense online customer reviews can be interpreted as post-consumption emotional traces, where users spontaneously articulate their affective responses to specific service attributes. From an Emotional Design perspective, these texts provide ecologically valid insights into how designed service elements are emotionally perceived in real-world contexts.

Different authors analyze online reviews to develop a data-driven faster, cost-effective and larger bigger-data scale alternative to methods traditionally used in Kansei Engineering (KE), an emotional design method, for identifying Kansei words, product features and the relationship between both (Chiu & Lin, 2018; Jiao & Qu, 2019; Li, Go & Jin, 2020; Lui, Cui & He, 2021). Usually, different Natural Language Processing (NLP) approaches are used to extract nouns in the reviews as product features and adjectives as Kansei words. Topic modelling with Latent Dirichlet Allocation (LDA), which refers to identifying latent topics (groups of words that appear related in the corpus of unstructured texts analyzed), is commonly applied to identify product attributes (Zhang et al., 2022; Lui, Cui & He, 2021; Ozyurt & Akcayol, 2021; Alzate, Arce-Urriza & Cebolla, 2022; Lui, Cui & He, 2021). It has also ben used dependency parsing to identify groups of related words that appear in the comments extracting from them product attributes and the evaluation (Jiao & Qu, 2019). However, KE uses a set of bipolar items or Kansei words (heavy-light, big-small) to assess customer emotional preferences of products instead of discrete emotions.

In this regards, even though the most common approach for assesing the emotional content of online review is Sentiment Analysis (SA), which classifies a comment as emotionally positive, negative, neutral or mixed (Kumar, Malik & Raman, 2022; Ozyurt & Akcayol, 2021; So, 2020; Rocklage, Rucker & Nordgren, 2021; Samuel, Rozzi & Palle, 2022) there is a broad consensus that SA has limitations for practical applications in the design of products and services, and that an analysis that considers the wide variety of emotions that humans can feel is necessary (Mao et al., 2019; Lelieveld & Hendriks 2021). Consequently, an increasing number of researchers follow this approach for emotional analysis of online reviews (Acheampong, Wenyu & Nunno-Mensah, 2020). However, there are very few published works applying the analysis of discrete emotions in online reviews for ED of products and services. Kumar and Ilavarasan (2020) propose differences in discrete emotions expressed in tweets before and after launching a product for guiding the process of new products development, but to authors knowledge, there are not published works addressing the relationship between topics and discrete emotions in online reviews.

In this sense, an approach gaining traction among researchers, is topic-sentiment combined analysis of online reviews with different purposes. For example, the Join Sentiment-Topic (JST) analysis that extracting different topics under the set of positive and negative reviews: positive and negative topics (Li,

Wu & Mai, 2019). Aspect-Based Sentiment Analysis (ABSA) assess the polarity of each attribute in each review and, finally aggregate the results to provide an overall sentiment for each aspect in the corpus of comments (Lui, Cui & He, 2021). But, to the authors knowledge, this is the first research exploring this approach for predicting discrete emotions.

In Sport, even though there is a growing body of published research showing a relevant role of emotions in realms such as Sports Marketing (Kwak, Kim & Hirt, 2011), Sports Management (Lin et al., 2022) or fan engagement (Billings, Butterworth & Turman, 2017; Aloufi & Saddik, 2018), as well as the relationship between sport products and services characteristics and customs emotions (Nusairat et al., 2020; Lin et al., 2022), ED is a rather new field of application. In this regard, though social media are acknowledged as an effective medium to observe the dynamics of emotions expressed by sports fans (Billings, Butterworth & Turman, 2017), emotional analysis of user online generated content is still very emerging and of our knowledge, research on data driven ED in sport has not been published before.

Jia (2019), for example, states that though exercisers are very active in posting online ratings and reviews regarding fitness clubs, research on perception and emotions has fundamentally relied on costly traditional ways of information gathering such as surveys and in-depth interviews.

Therefore, the purpose of this work is to analyze the relationship between sentiment and latent topics in customers online reviews with discrete emotions of fitness centers, as a way of offering finer grained information for product and services innovation, adopting a discrete emotion framework based on the assumption that emotions are qualitatively distinct affective states with specific experiential and functional properties. Discrete emotion models have shown particular relevance in service contexts, where specific emotions such as trust, anger or pride are directly linked to behavioral outcomes.

Method

Sample

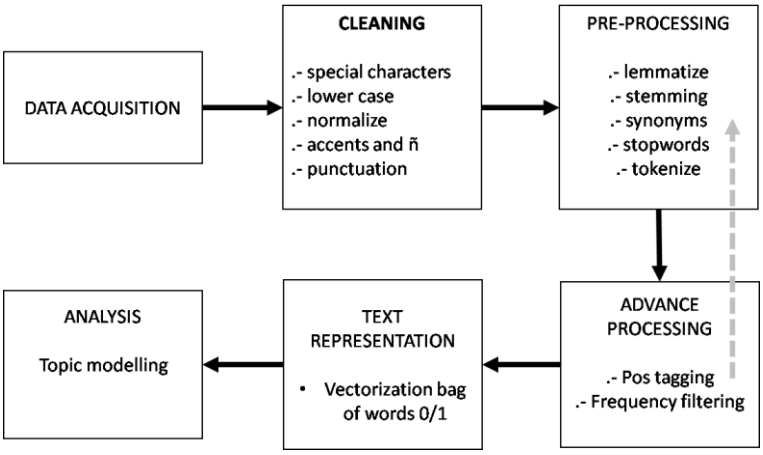
A total of 3,250 online reviews were collected from Google Maps review services corresponding to 38 fitness centers from Spain. Reviews included the text and the general evaluation in 5-stars ranking system (Mathayomchan & Taecharungroj, 2020). The number of comments collected for each fitness center ranged between 155 and 27 (M = 52.74, SD = 26.59).

Procedure

Comments were labelled in 10 different emotions by a panel of 5 volunteers and the most frequent ones were chosen. Following Štajner`s (2021) recommendations, cases with not a most frequent label were excluded. To make the labelling task easier, some emotion categories were double-labelled, and annotators were instructed to consider both when labelling a review. They were instructed to choose the emotion they think the person writing the review expresses. Emotions were selected from the literature starting by Plutchnik's anger, anticipation, disgust, fear, joy, sadness, surprise and trust, including love-pride as in Shaver's model (Shaver et al., 1987). In addition to classical discrete emotions, the category calm/relax was included to capture low-arousal positive affective states, which are especially relevant in wellness- and health-oriented services such as fitness centers.

The study follows a semi-automated process (Figure 1) for topic modelling of comments. Online review capturing was done manually, whereas all the text cleaning and processing was done automatically, except identification and reduction of synonyms, which required human intervention. The frequency filtering leaves only the words that appear in at least 3% of reviews and in no more than 30% of them. The upper threshold eliminates words that are too frequent because they refer to "structural" aspects of the comment set (for example "gym"). The lower limits reduce the sparseness of the resulting words-comments matrix. NLP tools in Python 3.0 (nltk and sklearn libraries) are used in a Jupyter notebook in an Anaconda environment.

Figure 1. Flow-chart of the process followed for topic modelling.



Source: Own creation

Data analysis

The analysis consisted first in topic modelling using LDA. Then, sentiment analysis was performed. Finally, the relationship between topics, sentiment and emotions was investigated following a machine learning approach.

Topic modelling using Latent Dirichlet Allocation

LDA was performed using the Gensim module in Python. The LDA model identifies keywords within the analyzed documents and proposes a distribution of topics in a randomly identified sample.

The following criteria were used to fix the number of topics to be extracted:

Criterion 1: Coherence

The coherence of the model refers to how different the words are in each topic. The more different they are, the more coherent the model is. It was measured using the C_V metric, which results from the application of a sliding window and segmentation of the top words in a set, as well as indirect confirmation based on the point-normalized mutual information and cosine similarity criteria. The model's hyperparameters, alpha and beta, related to the density of words in a text and in the set of texts respectively, were chosen according to maximize C_V.

Criterion 2: Low overlap measured by the distance between topics in the model

The greater the distance, the better. Jacckard distance was used to evaluate the overlap. It measures dissimilarity between sample sets. In this case, given topic 1 (t1) and topic 2 (t2), the Jackard distance (J) is computed as:

$$d_J(t1, t2) = (|t1 \cup t2| - |t1 \cap t2|)/(t1 \cup t2) \quad (1)$$

Where $|t1 \cup t2|$ stands for all the words that result from joining both topics, and $|t1 \cap t2|$ refers to words that are in both topics. In other words, it measures the ratio between words that do not overlap and the total of words in both topics.

Criterion 3: Interpretability of topics

LDA issues the weight of each word to each topic as well as the contribution of each topic to each comment. Topics are interpreted by considering words with higher contribution to them as well as reviews in which they have a high weight.

We used the tool pyLDAvis to interpret the topics. It extracts information from a fitted LDA topic model to inform an interactive web-based visualization (Sievert & Shirley, 2014). It includes a representation of the topics as circles in the two-dimensional plane whose centers are determined by computing the

distance between topics, and then by using multidimensional scaling to project the intertopic distances onto two dimensions. The area of each circle indicates the relative statistical weight of topics. It also displays a horizontal barchart representing the words that are the most useful for interpreting a selected topic. That barchart shows both the corpus-wide frequency of a word and its frequency in the topic selected. For each topic, words are sorted according to a relevance metric (λ) which combines two approaches to measure the association of a word to a topic: 1) a word as highly associated with a topic if its frequency in that topic is high (λ= 1) and 2) a word as highly associated with a topic if the "the ratio of a term's probability within a topic to its marginal probability across the corpus," is high (λ=0). Sievert and Shirley (2014) found that λ = 0.6 was optimal for interpretable results.

Since LDA is a probabilistic model, the averages of C_V and Jackard distance were calculated from five runs for an increasing number of topics from three to 15. The final solution should be a balance between low number of topics, high average coherence, high average distance and interpretability.

Sentiment Analysis

Sentiment analysis is done using textBlob from Natural Language ToolKit (NLTK). TextBlob returns polarity ranging between negative (-1) and positive (1) sentiments. Values lower than zero are classified as negative, equal to 0 as neutral and greater than 0 as positive.

Machine Learning Analysis

Data was split into test and train (60% and 40% respectively as data is unballanced). Different classification methods were applied to train a model on the train dataset which was tested afterwards on the test dataset. Cross validation for 20 random splits of data was implemented and the mean and standard deviation were computed for the metric wieghted f1-score from the confussion matrix (recommended for unballancede datasets, it is computed as:

f1_score = 2*(acccuray*sensitivity) / (accuracy+ sensitivity)) (2)

Values of f1 over 0.75 are usually recommended for a model to be good, obtained through Python 's scikit-learn module.

In the analysis, the target variable is each emotion transformed into a binary variable (0/ 1) being 1 that the text is labelled with this emotion. Prescriptors are new variables (stopics) computes as topic weight times polarity being -1 for negative sentiment and +1 for neutral and positive sentiments.

The method overally performing best for all emotions was chosen for a XAI analysis to identify the influence of each stopic in each emotion.

Explainable Artificial Intelligence (XAI): SHAP

The influence of predictors in elliciting or not a given emotion is measured using the Shap Library. It is model-agnostic and uses the Shapley values from game theory to estimate how each feature contributes to the prediction. We obtained shap_values to measure the importance of features and Shap summary plots to understand how each feature influences the predictions. Given the current set of feature values, the contribution of a feature value to the difference between the actual prediction and the mean prediction reflects the estimated Shap value. Features with positive Shapley Additive Explanations (SHAP) values positively impact the prediction, while those with negative values have a negative impact. The magnitude is a measure of how strong the effect is.

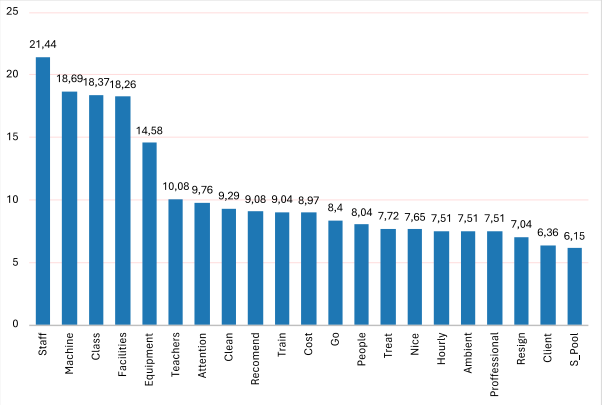
Summary plots allows understanding the type of influence combining the Shap value of the topic (red for high and blue for low) and the impact on model output (horizontal axis). In this way, a feature has a negative influence when high values (red dots) associate negative impact (horizontal axis) on the model output.

Results

Some topics show counterintuitive relationships between sentiment polarity and emotional outcomes, highlighting the distinction between evaluative valence and the specific emotion expressed by users.

The final dictionary after cleaning and pre-processing the comments included 91 words. Figure 2 shows the frequency distribution for the 20 most frequent words in the corpus of online reviews.

Figure 2. Frequency distribution (%) of words in the dictionary after cleaing and pre-processing.

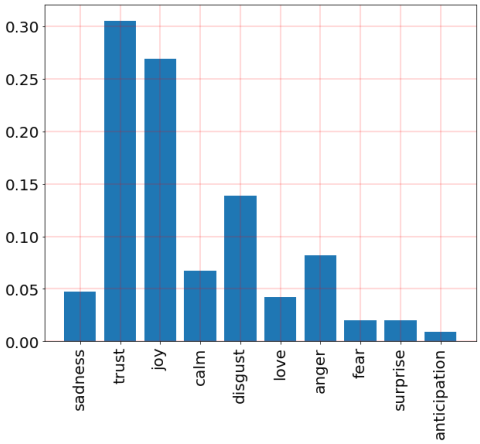


Source: Own creation

After Štajner`s (2021) recommendations, 150 (4.6%) cases with not a most frequent label were excluded. Thus, the dataset for the study consists of 3100 online reviews.

Figure 3 shows the frequency distribution of emotions for the labelled dataset of comments. Results confirm that the dataset is unbalanced. Trust is the most common emotion (30.5%) followed by joy (26.9%), disgust (13.8%) and anger (8.2%). The frequency of surprise (n=33) and anticipation (n=17) was very low and thus, were excluded from further analysis.

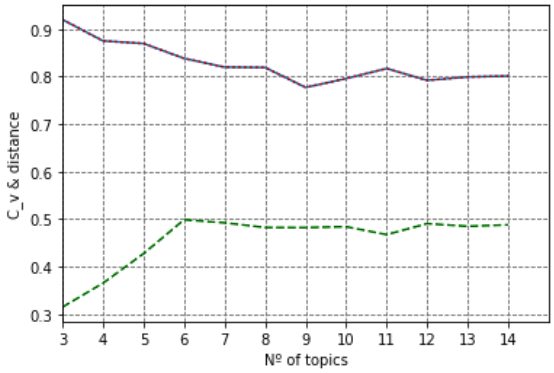
Figure 3. Frequency distribution (%) of emotions in the data set.



Source: Own creation.

Figure 4 shows the coherence (C_V) and distance between topics for the LDA model as a function of the number of topics. There are good solutions for both parameters between 6 and 12 topics.

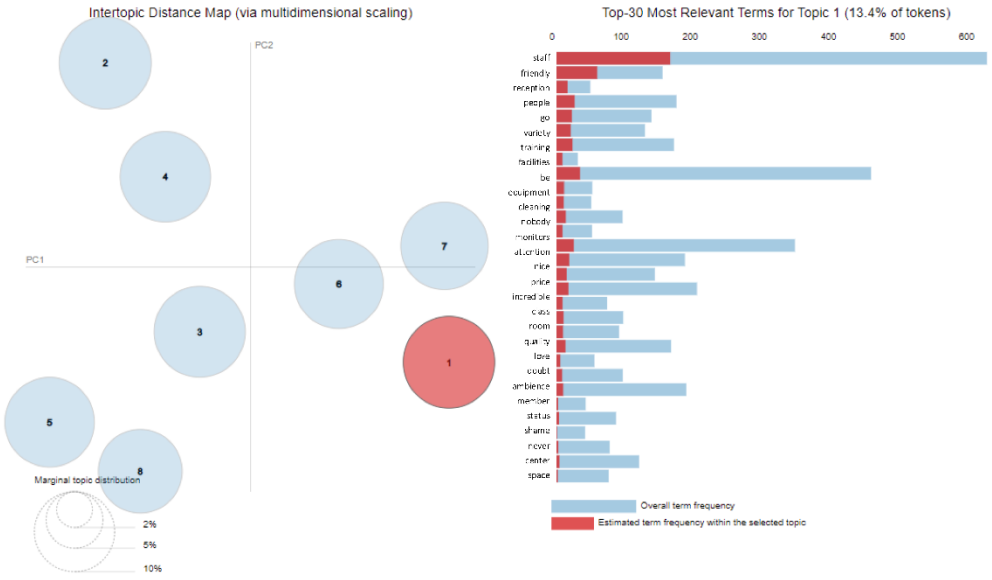
Figure 4. Coherence and distance between topics.



Source: Own creation.

After trying different options, a model with 8 topics was selected. Figure 5 shows Pyldavis visualization (λ= 0.6) in which topics do not overlap and all of them have a similar relevance (circle size).

Figure 5. Pyldavis result for 8 topics.



Source: Own creation.

Table 1 shows the 12 most relevant words for each topic and the label assignat to it.

Table 1. Topics showing the 12 most relevant words and label for each topic.

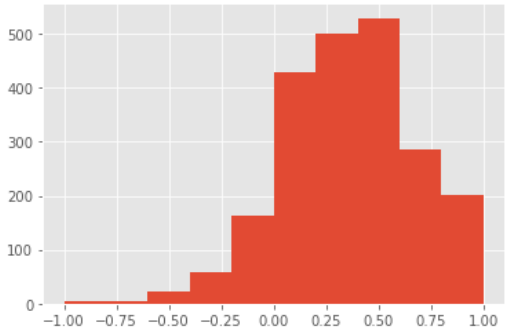
Label	Most relevant words											
T1. Unsubscribe during covid	0.07* covid	0.06* staff	0.05* can be	0.04* low	0.04* go	0.04* facilities	0.04* instructors	0.03* cost	0.03* instant	0.03* atention	003* impossi-ble	0.02* train
T2. Quality/ price	0.105* facilities	0.093* staff	0.064* cost	0.056* quality	0.049* professional	0.043* service	0.037* client	0.035* people	0.03* nice	0.026* train	0.024* condition	0.022* Instruc-tors
T3. Friendly atmosphere & staff, good price	0.12* staff	0.07* facilities	0.06* atmosphere	0.05* professional	0.04* friendly	0.04* cost	0.03* nice	0.03* instant	0.03* center	0.03* instruc-	0.02* last	0.02* variety

										tors		
T4. Unsubscribe, sauna, spool, quality	0.07* facilities	0.06* low	0.06* staff	0.04* atmosphere	0.04* instructors	0.03* pool	0.03* quality	0.03* sauna	0.03* exercise	0.02* train	0.02* cost	0.02* space
T5. Variety & friendly attention	0.177* staff	0.068* friendly	0.042* facilities	0.035* people	'0.033* instructors	0.031* train	0.030* go	0.029* variety	0.026* attention	0.025* cost	0.024* reception	0.022* nice
T6. mask, covid mesaures, crowded	0.154* instructors	0.072* facilities	0.056* staff	'0.040* attention	0.037* center	0.031* mask	0.029* go	0.027* people	0.023* train	0.021* atmos-phere	0.021* question	0.018* room
T7 All you need for training	0.082* facilities	0.056* staff	0.051* instructors	0.046* class	0.045* train	0.041* service	0.039* type	0.037* client	0.036* attention	0.032* question	0.029* covid	0.025* atmos-phere
T8. Cleanness & variety (lack of)	0.069* staff	0.059* zone	0.045* lack	0.044* pool	0.036* attention	0.036* room	0.033* facilities	0.032* variety	0.030* people	0.029* instruc-tors	0.028* client	0.028* cleaning

Note: Own creation
*Significant differences, p< .05

Results of sentiment analysis show that polarity distribution is skewed towards positive comments (Figure 6).

Figure 6. Results from sentiment analysis (polarity).



Source: Own creation.

Table 2. Mean f-1 weighted score (standard deviation) resulting from cross-validation for different classification methods for each emotion. Last row show the results for the fitted model.

Mean Method	Emotion (0/1)							
	Anger	Calm	Disgust	Fear	Joy	Love	Sadness	Trust
K-neighbours	0.89 (0.00)	0.94 (0.0)	0.81 (0.01)	0.97 (0.01)	0.54 (0.04)	0.94 (0.0)	0.91 (0.0)	0.62 (0.0)
Linear SVC	0.89 (0.01)	0.94 (0.00)	0.81 (0.01)	0.97 (0.01)	0.52 (0.04)	0.94 (0.0)	0.91 (0.0)	0.62 (0.1)
Log. Regression	0.88 (0.04)	0.62 (0.06)	0.78 (0.04)	0.85 (0.02)	0.5 (0.04)	0.67 (0.04)	0.73 (0.03)	0.48 (0.05)
Nearest centroid	0.89 (0.03)	0.25 (0.06)	0.84 (0.03)	0.92 (0.01)	0.38 (0.04)	0.46 (0.05)	0.86 (0.03)	0.31 (0.04)
Random Forest	0.9 (0.02)	0.94 (0.02)	0.83 (0.02)	0.96 (0.01)	0.61 (0.04)	0.94 (0.0)	0.9 (0.00)	0.64 (0.03)
XGB	0.9 (0.01)	0.94 (0.00)	0.83 (0.02)	0.96 (0.01)	0.61 (0.03)	0.94 (0.01)	0.9 (0.01)	0.65 (0.03)
XGB fitted	0.89	0.94	0.84	0.95	.61	0.94	.89	.89

Note: Own creation

Among the different ML methods, XGB shows a good performance for all emotions and SHAP was applied to it for each emotion. The last row of table 2 shows the weighted f1-score for the XGB model finally fitted and optimized for each emotion.

XGB is a GBT model. GBT is an ensemble method that performs classification by combining the outputs of individual trees. It combines many decision trees to reduce the risk of overfitting each individual tree. This process applies boosting, combining simple decision trees (weak learners) to improve against a loss function (Hastie, Tibshirani & Friedman, 2009). The analysis was performed via Python software using the Gradient Boosting Classifier from skicit learn (Pedregosa et al., 2011). In this study, the best combination of hyperparameters N_estimators (trees), Learning rate (how fast the model learns), and max depth of trees (how much a single tree can learn), was found using a grid search, the root mean squared error (RMSE) and accuracy with cross-validation. A common recommendation is to keep the learning rate low and increase the number of trees to reduce the risk of overfitting. The parameters chosen for the model were:

N_estimators = 100 trees; Learning rate = 0.25, and Max depth = None. (3)

Table 3 shows the results from SHAP. For better understanding the results, H+ refers to the impact of high values of topics (red dots in summary plots like the one showed in Figure 7) and H- to low values of topics. In this case, H+ are positive comments in which a topic has a moderate/ high weight whereas H- are negative comments in which a topic has a moderate/ high weight.

Table 3. Results from SHAP. Mean shap value and influence from summary plots.

Emotion	Impact	T1. Unsubscribe during covid	T2. Quality/ price	T3. Friendly atmosphere & staff, good price	T4. Unsubscribe, sauna, spool, quality	T5. Variety & friendly attention	T6. mask, covid mesaures, crowded	T7 All you need for training	T8. Cleanness & variety
anger	SHAP	.38	.19	.83	.4	.41	.29	0,3	.42
	H+	inc	inc	dec	dec	dec	dec	inc	Dec
	H-	dec	dec	inc	inc	inc	inc	Dec	inc
calm	SHAP	0.44	0.4	0.15	0.26	0.27	0.66	0.58	0.47
	H+	Inc	inc	Inc/ dec	inc	Inc	dec	inc	inc
	H-	Dec	Dec	inc	dec	dec	inc	dec	dec
disgust	SHAP	.36	.28	.37	.46	.29	.21	.52	.58
	H+	inc	dec	dec	Inc	dec	Dec	dec	inc
	H-	dec	Inc/dec	dec	dec	Inc	Inc	Inc	dec
Fear	SHAP	.18	.27	.16	.52	.1	.65	.57	.16
	H+	Inc	dec	Inc	Dec	Nef	Dec	Inc	inc
	H-	dec	inc	dec	inc		Low inc	dec	dec
joy	SHAP	.32	.36	.56	.42	.26	.31	.24	.40
	H+	Dec	Inc	Inc	Inc	Inc	Inc	Inc	Inc
	H-	inc	L dec	dec	dec	dec	Dec	dec	ded
love	SHAP	.31	,37	.61	.22	.73	.39	.46	.34
	H+	Dec	Dec/ inc	Inc	inc	Inc	Dec	Dec	Inc
	H-	inc		M dec		dec	inc		ne
sadness	SHAP	.67	.58	.42	.34	.43	.67	.6	.31
	H+	Inc	Dec	Dec	Dec	Inc	Dec	inc	--
	H-	Dec	inc	inc	inc	dec	inc	Dec	--
trust	SHAP	.36	.69	.72	.29	.43	.59	.57	.54
	H+	Inc	Inc	Inc	inc	Inc	Dec	Dec	Inc
	H-		Dec	Dec	Dec	Dec	Dec	Inc	Dec

Note: H+ = impact of high value on positive feedback (red dots); H- = impact of high value on negative feedback (blue dots); dis = decrease; inc = increase; Own creation

Following, the results for each emotion are described:

Anger. T3. Friendly atmosphere and staff are the most influencing topic in a way that positive comments with a high weight of T3 reduce anger and vice versa. T1. Unsubscribe during covid; T4. Unsubscribe, sauna, spool, quality, T5. Variety and friendly attention and T8. Cleanness and variety (lack of) follow with a similar level of influence. Positive Comments with a high or moderate weight of T4, T5 and T8 reduce anger, and vice versa. Positive Comments with a high or moderate weight of T1 increase anger and vice versa.

Calm. T6 mask, covid measures, crowded; shows the highest influence in calm, followed by T7, all you need for training. Positive comments with high weight of T6 decrease calm and vice versa, whereas positive comments with moderate- to- high weight of T7 increase calm and vice versa.

Disgust. T8. Cleanness and variety (lack of) shows the highest influence in disgust, followed by T7 All you need for training and T4. Unsubscribe, sauna, spool, quality. Positive comments with a high weight of T8 increase disgust and vice versa, positive ones with high weight of T7 decrease disgust and positive comments with high weight of t4 increase disgust and vice versa.

Fear. T6. mask, covid measures, crowded shows the highest influence in fear, followed by T7 All you need for training and T4, Unsubscribe, sauna, spool, quality. Positive comments with high weight of T6 decrease fear and vice versa. High weight of T7 in positive comments increase fear and vice versa, whereas positive comments with high weight of T4 decrease fear and vice versa.

Joy/satisfaction. T3. Friendly atmosphere & staff, good price shows the highest influence in joy followed by T4, Unsubscribe, sauna, spool, quality. Positive comments with high weight of T3 moderately increase it and vice versa low. Also, positive comments with high weight of T4 increase joy and vice versa.

Love/pride. T5 Variety and friendly attention show the highest influence in love followed by T4, Unsubscribe, sauna, spool, quality. Positive comments with high weight of T5 and T4 increase it and vice versa.

Sadness/deception. T1 Unsubscribe during covid and T6 mask, covid measures, crowded, show the highest influence in deception followed by T7 All you need for training and T2, Quality/ price. Positive comments with high weight of T1 increase it and vice versa. Positive comments with high weight of T6,

T7 and T2 decrease it and vice versa.

Trust. T3. Friendly atmosphere and staff, good show the highest influence in price followed by T2, Quality/ price, T6. mask, covid measures, crowded, T7 All you need for training, T8. Cleanness and variety (lack of). Positive comments with high weight of T3, T2, T6 and T8 increase trust and vice versa. Negative comments with high weight of T7 increase it and vice versa.

Discussion

The present research establishes the influence of different attributes of fitness centers in the emotions expressed online by customers. The topics identified mostly coincide with those found by a study that analyzed influence of online reviews on star rating (Jia, 2019). The author showed that general comment, environment, reception, responsibility, lose weight, motivation, location and course influence in obtaining 5 stars, whereas sales, major brand, private coach and shower, associate with 3 stars. Regarding attributes influence in emotion, results are not easily compared with published research. Up to date, researchers have shown the relationship of different attributes of fitness centers such as indoor environmental quality, with attitudes (Nusairat et al., 2020), perceived service quality and perceived psychological safety (Lin et al., 2022) but not with emotions. Also, in general, to date, published research study the relationship between topics/ product attributes and different customer outcomes such as satisfaction (So, 2020; Wang, Lu & Tan, 2018) or sales (Li, Wu & Mai, 2019), but to authors knowledge, not with discrete emotions.

Additionally, though there is a growing number of researchers combining sentiment and topics analysis with different purposes, to authors knowledge, this is the first research exploring this approach for predicting discrete emotions. Besides, approaches applied in literature usually assign to each topic an aggregated sentiment along a set of comments. For example, ABSA extracts the aspects or features of a product or service, to assign a sentiment polarity to each aspect based on the context of each sentence and, finally aggregate the results to provide an overall sentiment for each aspect in the corpus of comments (Ozyurt & Akcayol, 2021; Lui, Cui & He, 2021; Chen et al., 2022; Zhai et al., 2022; Wan et al., 2020; Zhang et al., 2022; Xiang et al., 2023). But the approach followed in the present research is using as predictors the result of multiplying the weight of the topic in each comment by the polarity of the comment's sentiment (+1, -1) which authors name as "polarized topics" or topics.

In this regard, results show that "topics" generate ML models with good to excellent (f1-score > 0.84) performance for all emotions but for joy/ satisfaction (f1 = 0.61). This result could be due to a more

variety in the criteria (less consensus) annotators consider when assigning joy to a comment. Also, results from cross-validation (table 2) show a lower average performance (0.62) for trust. Confusion between joy and trust could be a source of variability leading to poorer performance of ML models for these emotions. This points towards the complexity to distinguish between some emotional categories as a shortcoming of manual labelling and to the need for future research investigating this topic as discussed below.

On this regard, the choice of emotional categories is a topic for discussion. Existing models for establishing the emotional labels can be grouped into discrete or dimensional (Wang et al., 2012; Seyeditabari et al., 2019; Kumar, Malik & Raman, 2022; So, 2020; Wani et al., 2019). Dimensional models use different properties to define emotions, like for example, valence and arousal of Circumplex Model (Russel, 1980; Mao et al., 2019) or Pleasure, Arousal and Dominance of Emotional State Model (Yi et al., 2023; Shah Abbasi & Yan, 2023). Some authors recommend dimensional models to account for interdependencies and similarities between emotions that make difficult to distinguish between categories like, for example, relax and calm (Acheampong, Wenyu & Nunno-Mensah, 2020). Others (Gupta & Srinivasan, 2020) consider that in models like the Circumplex model few emotions are clustered so close to each other that, there is a high probability of mislabelling them. On the other hand, it has been observed that discrete emotion analysis clearly outperforms central affect models (valence and arousal) when studying firm–customer encounters (Kranzbühler et al., 2020). In this context, generally, discrete emotion models have been widely adopted due to its simplicity.

In this sense, the present research combines discrete emotions models most applied in literature: Shaver (Shaver et al., 1987); Ekman's basic emotions (Ekman, 1992), and Plutchnik (Plutchik & Conte, 1997) adding calm to account for low activation positive emotions that authors expected to appear in reviews about services. Actually, almost 7% of comments were labelled as expressing calm. Also, double labelling was included to make easier annotators task accounting for interdependencies and similarities between emotions that make difficult to distinguish between categories like, for example, relax and calm as suggested by some authors (Acheampong, Wenyu & Nunno-Mensah, 2020).

The results show that, as reported by different authors, the best performing model is achieved by complex black-box like method, such as ensemble Gradient Boosting Trees (GBT) (So, 2020; Joung & Kim 2021). Various Explainable Artificial Intelligence (XAI) methods have been developed to interpret the predictions of complex models and explain them to humans in an understandable way (Murdoch et al., 2019). SHAP is one of the most widely used (Lundberg and Lee 2017; Lundberg et al. 2018; Joung & Kim 2021). The application of SHAP to these models allows understanding the way each "polarized topic" influences each emotion.

In this line, Joung and Kim (2021) combined LDA to detect product attributes from online reviews with ABSA to determine the importance of the identified attributes. They evaluated the impact of the sentiments of each product attribute on the overall rating via an explainable, deep neural network applying SHAP on a rather large dataset of 33,379 cell phones reviews. Their results outperformed previous sentiment analyses and neural network-based methods. However, to authors knowledge, no publications have used applications combining XAI and ML classification methods like, for example, XGB, to establish the influence of product attributes in emotions.

In this case, the main findings of the present study show that the influence of topics (T) in emotions is as follows:

T1. Unsubscribe during covid. This is a main topic identified pointing towards a specific issue that took place during covid-19 close-down of gyms. Weighting in positive/ neutral comments increases anger, calm and sadness/ deception, and vice versa.

T2. Quality/ price**.** Weighting in positive/ neutral comments increases calm, trust and decreases sadness/ deception; and vice versa. This concept was described in a published work on hotel services in relation between quality and price as a hotel utility (Xu, 2019).

T3. Friendly atmosphere and staff, good price. Weighting in positive/ neutral comments decrease anger and disgust, increase joy/ satisfaction, love/ pride and trust, and vice versa. Atmosphere, ambience and staff are commonly cited as relevant in different works about services in general (Ogruk, Anderson &

Nacass, 2018) and fitness centers in particular (Jia, 2019; Nusairat et al., 2020). It has been demonstrated that the proper blend of ambience variables triggers a customer's positive emotional state leading to an enjoyable, memorable store visit, more time and money spent in the store, and more products purchased (Ogruk, Anderson & Nacass, 2018). It has also been remarked the importance of consumers' interactions with employees as a trigger and moderator of consumers' emotions and their influence on evaluations as well as purchase and sharing behavior (Kranzbühler et al., 2020).

T4. Unsubscribe, Sauna, spool, quality. Weighting in positive/ neutral comments decreases anger, fear and increases disgust and joy.

T5. Variety and friendly attention. Weighting in positive/ neutral comments decreases anger and increases love/ pride, and vice versa. This is complementary with T3 by making the difference between staff and supporting staff in fitness centers, coincident to that found in other work that identifies a similar topic including reception and Sales (Jia, 2019). Interestingly both have a similar influence in anger but T5 only influences in love/ pride.

T6. Mask, covid measures, crowded. Weighting in positive/ neutral comments decreases calm, fear, sadness/ deception and trust, and vice versa.

T7. All you need for training. Weighting in positive/ neutral comments increases calm, fear and sadness/ deception, and decreases disgust/ dislike and trust. A topic called general comments has been reported in another publication that expresses exercisers' general estimation and willingness to return or make recommendations (Jia, 2019).

T8. Cleanness and Variety. Weighting in positive/ neutral comments decreases anger and increases calm, disgust, joy and trust; and vice versa.

These results may have practical implications for fitness centers managers and professionals. However, there seems to be some contradictory results regarding the polarity of topics and valence of associated emotions. For example, T1 (Unsubscribe during covid) weighting in positive/ neutral comments provokes anger and sadness/deception, which are not positive emotions. Similarly, T4 (Unsubscribe, Sauna, spool, quality) weighting in positive/ neutral comments increases disgust and joy at the same time; whereas T6 (Mask, covid measures, crowded) weighting in positive/ neutral comments decreases calm and trust. Also, T7 (All you need for training) weighting in positive/ neutral comments increases fear and sadness/ deception and decreases trust.

We used TextBlob for sentiment analysis. It follows a lexicon-based approach using a pre-defined dictionary classifying negative and positive words. After assigning individual scores to all the words in a comment, final sentiment is calculated by averaging all the sentiments. But, for emotion labelling, annotators may consider the appearance of a single anchorage word (e.g. "I am happy") or a more complex analysis of the comment considering semantic relationships. For example, comments may have more than one sentence with different polarities each (e.g. "This gym is a good place for training, but it is always crowded") in a way that the polarity of the sentence would tend to be neutral whereas we may perceive that the user is expressing deception. This opens an interesting line for future research.

For assigning emotional labels to texts, the most common approach is manual annotation. Though it is alleged to be more accurate, manual labelling is time consuming and costly (Joung & Kim, 2021) as well as a complex task in a way that it is not easy that human annotators reach a consensus. Some 150 comments (4.6%) were excluded from the present study because do not having a most frequent label. On this regard, there is a big interest in developing automatic or semiautomatic (Tafreshi & Diab, 2018; Janssens et al., 2014) labelling approaches. Weak labelling approaches (Janssens et al., 2014), for example, use twits hashtags or emotional seed words, together with some manual labels, as weak and noisy initial emotional labels for an automatic labelling process. As a result of this intense research, there are available some open codes and manually labelled datasets, as well as tools, for emotion recognition in texts (Acheampong, Wenyu & Nunno-Mensah, 2020), but these tools and models show poor performance outside the domain and context of the dataset used for training them (Joung & Kim, 2021; Štajner, 2021; Acheampong, Wenyu & Nunno-Mensah, 2020, Jiao & Qu, 2019). Future research shold focus in automatic and weak emotional labelling of online reviews of sport services like fitness centers.

Also, emotional design traditionally attempts to relate product design attributes with emotions, but the approach followed in the present research considers customers online reviews in which they comment

on certain product/ service elements they consider relevant for the issue they are commenting on but may leave out product attributes relevant for the emotional experience. In this sense, it would be interesting a combined analysis considering also descriptive attributes of the service that may be obtained when gathering online comments (price, location, etc.).

With respect to extrapolating results to other services and products, some authors (Jiao & Qu, 2019) state that applicability of online reviews for emotional design is limited and depends on the type of product.

Conclusions

Results of the present research show the relationship of discrete emotions with fitness center attributes offering managers and professionals valuable knowledge. Positive comments with high weight of T1. Unsubscribe during covid increase anger and reduce sadness/deception. Positive comments with high weight of T2. Quality/price decrease sadness/deception and increase trust. Positive comments with weight of T3. Friendly atmosphere and staff reduce anger, increase joy/satisfaction, love and trust. Positive comments with weight of T4. Unsubscribe, sauna, spool, quality reduce anger and fear and increase disgust/dislike and joy/ satisfaction. Positive comments with high weight of T5. Variety and friendly attention decrease anger and increase love/pride. Positive comments with high weight of T6. mask, covid measures, crowded decrease calm/relax, fear, sadness/deception and trust. Positive comments with high weight of T7. All you need for training increase calm/relax, fear and sadness/deception, whereas decrease disgust/dislike and trust. Positive comments with high weight of T8. Cleanness and variety decrease anger and increase disgust/dislike and trust.

This study shows that combining LDA with sentiment analysis of online reviews to build complex ML models and using SHAP to explain them, can be a time, cost and scale effective alternative for emotional design of product and services. The outcomes of this study suggest new channels for future research involving the application of XAI in product development and emotional engineering.

The results can be of use to sport managers, for fitness centers managers, since they may know which attributes and in which direction, to work in to elicit and avoid certain emotions. For example, they can work on generating positive comments with weight of T3. Friendly atmosphere and staff, good price to decrease anger and disgust, and increasing joy/ satisfaction, love/ pride and trust, or to put the focus in T5. Variety and friendly attention to decrease anger and increase love/ pride.

This study contributes by proposing a novel framework that combines topic modelling, sentiment analysis and explainable machine learning to predict discrete emotions from online reviews, providing both theoretical and practical implications for emotional design in service contexts.

Acknowledgments

IGOID-SPORTECH SL, Universidad de Castilla-La Mancha, Toledo, Spain.

DRQ Data Driver, Valencia, Spain.

Department of Sports Sciences, Faculty of Medicine, Health and Sports, Universidad Europea de Madrid, Villaviciosa de Odón, Madrid, Spain.

Real Madrid Graduate School, Faculty of Medicine, Health and Sports, Universidad Europea de Madrid, Madrid, Spain.

iGEO, School of Physical Education, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile.

Metis Research Lab, Facultad de Negocios y Tecnología, Universidad Alfonso X el Sabio (UAX), Madrid, Spain.

Financing

This article has been supported by grant PTQ2021-012032 funded by MICIU/AEI/10.13039/501100011033.

References

Acheampong, F. A., Wenyu, C., & Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2(7), e12189 https://doi.org/10.1002/eng2.12189

Aloufi, S. & Saddik, A.E. (2018). Sentiment Identification in Football-Specific Tweets. *IEEE Access*, 6, 78609-78621. https://doi.org/10.1109/ACCESS.2018.2885117

Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. *Journal of Retailing and Consumer Services*, 67, 102989. https://doi.org/10.1016/j.jretconser.2022.102989

Billings, A. C., Butterworth, M. L., & Turman, P. D. (2017). *Communication and sport: Surveying the field*. Sage publications.

Chaklader, R., & Parkinson, M. B. (2017). Data-driven sizing specification utilizing consumer text reviews. *Journal of Mechanical Design*, *139*(11), 111406. https://doi.org/10.1115/1.4037476

Chen, X., Sun, S., Zhang, Z., Ma, Z., Wu, X., Li, H., ... & Zhang, K. (2022). Consumer shopping emotion and interest database: a unique database with a multimodal emotion recognition method for retail service robots to infer consumer shopping intentions better than humans. *Journal of Electronic Imaging*, *31*(6), 061807-061807. https://doi.org/10.1117/1.JEI.31.6.061807

Chiu, M. C., & Lin, K. Z. (2018). Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. *Advanced Engineering Informatics*, 38, 826-839. https://doi.org/10.1016/j.aei.2018.11.002

Fan, W., Liu, Y., Li, H., Tuunainen, V. K., & Lin, Y. (2022). Quantifying the effects of online review content structures on hotel review helpfulness. *Internet Research*, *32*(7), 202-227. https://doi.org/10.1108/INTR-11-2019-0452

Fernandes, S., Panda, R., Venkatesh, V. G., Swar, B. N., & Shi, Y. (2022). Measuring the impact of online reviews on consumer purchase decisions–A scale development study. *Journal of Retailing and Consumer Services*, *68*, 103066. https://doi.org/10.1016/j.jretconser.2022.103066

Guo, F., Cao, Y., Ding, Y., Liu, W., & Zhang, X. (2014). A multimodal measurement method of users' emotional experiences shopping online. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *25*(5), 585-598. https://doi.org/10.1002/hfm.20577

Guo, J., Wang, X., & Wu, Y. (2020). Positive emotion bias: Role of emotional content from online customer reviews in purchase decisions. *Journal of Retailing and Consumer Services*, 52, 101891. https://doi.org/10.1016/j.jretconser.2019.101891

Hastie, T., Tibshirani, R., Friedman, J. H. (2009). "10. Boosting and Additive Trees". *The Elements of Statistical Learning* (2nd ed.). New York: Springer. pp. 337–384. ISBN 978-0-387-84857-0. https://doi.org/10.1007/978-0-387-84858-7_10

Janssens, O., Verstockt, S., Mannens, E., Van Hoecke, S., & Van de Walle, R. (2014). Influence of weak labels for emotion recognition of tweets. In *Mining Intelligence and Knowledge Exploration: Second International Conference, MIKE 2014, Cork, Ireland, December 10-12, 2014. Proceedings* (pp. 108-118). Springer International Publishing. https://doi.org/10.1007/978-3-319-13817-6_12

Jia, S. (2019). Toward a better fitness club: Evidence from exerciser online rating and review using latent Dirichlet allocation and support vector machine. *International Journal of Market Research*, *61*(1), 64-76. https://doi.org/10.1177/1470785318770571

Jiao, Y., & Qu, Q. X. (2019). A proposal for Kansei knowledge extraction method based on natural language processing technology and online product reviews. *Computers in Industry*, 108, 1-11. https://doi.org/10.1016/j.compind.2019.02.011

Kim, J., & Gupta, P. (2012). Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research*, *65*(7), 985-992. https://doi.org/10.1016/j.jbusres.2011.04.013

Kranzbühler, A. M., Zerres, A., Kleijnen, M. H., & Verlegh, P. W. (2020). Beyond valence: A meta-analysis of discrete emotions in firm-customer encounters. *Journal of the Academy of Marketing Science*, *48*, 478-498. https://doi.org/10.1007/s11747-019-00707-0

Kumar, P., Malik, S., & Raman, B. (2022). Hybrid fusion based interpretable multimodal emotion recognition with insufficient labelled data. arXiv preprint arXiv:2208.11450. https://doi.org/10.48550/arXiv.2208.11450

Kwak, D. , Kim, Y. & Hirt, E. (2011). Exploring the role of emotions on sport consumers' behavioral and cognitive responses to marketing stimuli. *European Sport management quarterly*, *11*(3), 225-250. https://doi.org/10.1080/16184742.2011.577792

Lelieveld, G. J., & Hendriks, H. (2021). The interpersonal effects of distinct emotions in online reviews. *Cognition and Emotion*, *35*(7), 1257-1280. https://doi.org/10.1080/02699931.2021.1947199

Li, L., Goh, T. T., & Jin, D. (2020). How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. *Neural Computing and Applications*, *32*, 4387-4415. https://doi.org/10.1007/s00521-018-3865-7

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management*, 56(2), 172-184. https://doi.org/10.1016/j.im.2018.04.007

Lin, P. T., Vu, T. T., Nguyen, V. P., & Wu, Q. (2022). Self-determination theory and accountant employees' psychological wellbeing: The roles of positive affectivity and psychological safety. *Frontiers in Psychology*, *13*, 870771. https://doi.org/10.3389/fpsyg.2022.870771

Liu, H., Cui, T., & He, M. (2021). Product optimization design based on online review and orthogonal experiment under the background of big data. Proceedings of the Institution of Mechanical Engineers, *Part E: Journal of Process Mechanical Engineering*, 235(1), 52-65. https://doi.org/10.1177/0954408920943690

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., ... & Lee, S. I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering*, *2*(10), 749-760. https://doi.org/10.1038/s41551-018-0304-0

Mao, Y., Dai, C., Zhang, Y., Shen, S., Ma, H., & Chen, R. (2019). Impact of Review Emotions on Sales: The Moderating Role of Product Type. *Review of Integrative Business and Economics Research*, 8(4), 97.

Mathayomchan, B., & Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. *International Journal of Hospitality Management*, *90*, 102641. https://doi.org/10.1016/j.ijhm.2020.102641

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Interpretable machine learning: definitions, methods, and applications. *arXiv preprint arXiv:1901.04592*. https://doi.org/10.48550/arXiv.1901.04592

Nusairat, N. & Hammouri, Q., Al-Ghadir, H., Ahmad, A. & Abuhashesh, M. (2020). Fitness centers ambience-customer behavioral intentions relationship: The mediating role of customer emotional states. *International Journal of Business and Management*, 15(9), 93. https://doi.org/10.5539/ijbm.v15n9p93

Ozyurt, B., & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with Applications*, 168, 114231. https://doi.org/10.1016/j.eswa.2020.114231

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830. http://jmlr.org/papers/v12/pedregosa11a.html

Plutchik, R., & Conte, H. (1997). *Circumplex models of personality and emotions*. American Psychological Association.

Rathore, A. K., & Ilavarasan, P. V. (2020). Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products. *International Journal of Information Management*, *50*, 111-127. https://doi.org/10.1016/j.ijinfomgt.2019.05.015

Rocklage, M., Rucker, D., & Nordgren, L. (2021). Mass-scale emotionality reveals human behaviour and marketplace success. *Nature human behaviour*, 5(10), 1323-1329. https://doi.org/10.1038/s41562-021-01098-5

Russel, J.A.: A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, 39(6), 1161-1178, 1980. https://doi.org/10.1037/h0077714

Samuel, J., Rozzi, G., & Palle, R. (2022). The dark side of sentiment analysis: An exploratory review using lexicons, dictionaries, and a statistical monkey and chimp. Dictionaries, and a Statistical Monkey and Chimp (January 6, 2022). http://dx.doi.org/10.2139/ssrn.4000087

Seyeditabari, A., Tabari, N., Gholizadeh, S., & Zadrozny, W. (2019). Emotion detection in text: focusing on latent representation. arXiv preprint arXiv:1907.09369. https://doi.org/10.48550/arXiv.1907.09369

Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63-70). https://aclanthology.org/W14-3110.pdf

Shah, A. M., Abbasi, A. Z., & Yan, X. (2023). Do online peer reviews stimulate diners' continued log-in behavior: Investigating the role of emotions in the O2O meal delivery apps context. *Journal of Retailing and Consumer Services*, *72*, 103234. https://doi.org/10.1016/j.jretconser.2022.103234

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52(6), 1061–1086. https://doi.org/10.1037/0022-3514.52.6.1061

So, C. (2020). Understanding the prediction mechanism of sentiments by XAI visualization. In *Proceedings of the 4th international conference on natural language processing and information retrieval* (75-80). https://doi.org/10.1145/3443279.3443284

Štajner, S. (2021, september). *Exploring Reliability of Gold Labels for Emotion Detection in Twitter.* [Paper] International Conference on Recent Advances in Natural Language Processing, 1350-1359 Online. https://doi.org/10.26615/978-954-452-072-4_151

Tafreshi, S., & Diab, M. (2018). Emotion Detection and Classification in a Multigenre Corpus with Joint Multi-Task Deep Learning. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2905–2913, Santa Fe, New Mexico, USA. Association for Computational Linguistics. https://aclanthology.org/C18-1246

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012, 3-5 september). *Harnessing twitter" big data" for automatic emotion identification.* [Proceeding] 2012 *International Conference on Privacy, Security, Risk and Trust and 2012 International Conferenece on Social Computing* (pp. 587-592), Amsterdam Netherlands. https://doi.org/10.1109/PASSAT-SocialCom30175.2012

Wang, Y., Lu, X., & Tan, Y. (2018). Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electronic commerce research and applications,* 29, 1-11. https://doi.org/10.1016/j.elerap.2018.03.003

Wani, M. A., Bours, P., Agarwal, N., & Jabin, S. (2019). Emotion-based mining for gender prediction in online social networks. In *Proceedings of the ACM, International Conference on Machine Learning and Data Science*.

Xiang, M., Zhong, D., Han, M., & Lv, K. (2023, July). A Study on Online Health Community Users' Information Demands Based on the BERT-LDA Model. *Healthcare,* 11(15), 2142. https://doi.org/10.3390/healthcare11152142

Xu, X. (2019). Examining the relevance of online customer textual reviews on hotels' product and service attributes. *Journal of Hospitality & Tourism Research*, 43(1), 141-163. https://doi.org/10.1177/1096348018764573

Yi, Q., Khan, J., Su, Y., Tong, J., & Zhao, S. (2023). Impulse buying tendency in live-stream commerce: The role of viewing frequency and anticipated emotions influencing scarcity-induced purchase decision. *Journal of Retailing and Consumer Services*, 75, 103534.https://doi.org/10.1016/j.jretconser.2023.103534.

Zhai, Z., Chen, H., Feng, F., Li, R., & Wang, X. (2022, December). COM-MRC: A COntext-masked machine reading comprehension framework for aspect sentiment triplet extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 3230-3241). https://doi.org/10.18653/v1/2022.emnlp-main.212

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11), 1101911038.https://doi.org/10.1109/TKDE.2022.3230975

Authors' and translators' details:

Enrique Alcántara Alcover	enrique.alcantara@igoidsportec.com	Author
Antonio Hernández Martín	antonio.hernandez@universidadeuropea.es	Author
Ricardo Cuevas Campos	ricardo.cuevas@uclm.es	Author
Daniel Duclos Bastías	daniel.duclos@pucv.cl	Author