International Journal of Multicriteria Decision Making

ISSN online: 2040-1078 - ISSN print: 2040-106X
https://www.inderscience.com/ijmcdm

Ranking based on user comments with multiple criteria: the case of Greek restaurants in Athens

Dimitrios Novas, Dimitrios Papakyriakopoulos, Elizabeth Powlesland-Kartaloglou, Anastasia Griva

DOI: 10.1504/IJMCDM.2023.10058937

Article History:

Received:	18 November 2021
Last revised:	05 December 2022
Accepted:	20 January 2023
Published online:	11 October 2023

Ranking based on user comments with multiple criteria: the case of Greek restaurants in Athens

Dimitrios Novas*,
Dimitrios Papakyriakopoulos and
Elizabeth Powlesland-Kartaloglou

West Attica University,
Egaleo City, 250 Thivon & P. Ralli Str., Egaleo,
Postal Code 12241, Athens, Greece
Email: dnovas@uniwa.gr
Email: dpapak@uniwa.gr
Email: elkarta@uniwa.gr
*Corresponding author

Anastasia Griva

Lero – The Science Foundation Ireland Research Centre for Software,
J.E. Cairnes School of Business and Economics,
University of Galway,
Upper Newcastle Road, H91wn80, Galway, Ireland
Email: anastasia.griva@universityofgalway.ie

Abstract: User generated content (UGC) is a valuable resource for the multi-criteria decision-making process. Platforms, such as TripAdvisor, enable the registration of comments and post rankings. However, existing ranking mechanisms are not transparent, making it impossible to evaluate their integrity and accuracy. Additionally, it has been shown that these rankings are, in some cases, inaccurate and misleading. This prompted the need for developing a ranking mechanism based on UGC to capture experiences and feelings, rather than quantitative data (bubbles/stars). The proposed ranking mechanism employs LDA to generate topics, which are transformed with fuzzy logic to variables, used to produce ranking results. We empirically studied the proposed ranking mechanism using TripAdvisor's user comments on a sample of restaurants located in Athens, Greece and compared the results to a simple quantitative ranking scheme. In some cases, the ranking differed. Further investigation is needed to address the limitations encountered in this research.

Keywords: multi-criteria analysis; text mining; TripAdvisor; ranking; fuzzy numbers; restaurants.

Reference to this paper should be made as follows: Novas, D., Papakyriakopoulos, D., Powlesland-Kartaloglou, E. and Griva, A. (2023) 'Ranking based on user comments with multiple criteria: the case of Greek restaurants in Athens', *Int. J. Multicriteria Decision Making*, Vol. 9, No. 3, pp.204–230.

Biographical notes: Dimitrios Novas is a Lecturer at Business Administration Department of West Attica University. He holds a BSc in Mathematics from University of Athens, MSc in DSS from University of Piraeus and MSc in

Mathematics from the Greek Open University. He has worked as a Production Manager in the media sector and teaches quantitative methods for the last 20 years. His research interests include fuzzy logic, decision support systems and ranking methods. His research has been presented in national and international conferences.

Dimitrios Papakyriakopoulos is a Lecturer at Business Administration Department of West Attica University. He holds a BSc in Informatics, MSc in Information Systems from Athens University of Economics and Business and PhD in Information Systems and Supply Chain Management also from Athens University of Economics and Business. He has worked as a researcher in various EU-funded projects and as special advisor in e-government for the Greek government. His research areas are business analytics, machine learning and data science. He has published in several scientific journals including *Decision Support Systems*, *Expert Systems with Applications*, *Electronic Markets*, etc.

Elizabeth Powlesland-Kartaloglou is a Lab Teaching Personnel at Business Administration Department of West Attica University. She holds a BSc in Business Management from Technological Institution of Athens, Greece and BSc in Computer Science from Rhode Island College, RI-USA. She also holds an MSc in Information Systems from Linnaeus University, Sweden. She has worked as a Technical Computer Analyst in RI-USA and has an extensive teaching experience in various academic fields in business administration. Her research interests include qualitative research methods in information systems and content analysis.

Anastasia Griva is a Postdoctoral Research Fellow at Lero | The Science Foundation Ireland Software Research Centre at NUI Galway. She received her PhD in Business Analytics from Athens University of Economics and Business, in Greece. Her research interests lie in the areas of business analytics for decision support, artificial intelligence, and customer behaviour. She has published in journals such as *Expert Systems with Applications*, *IEEE Software*, and *Information Systems Frontiers*. Her research has been also presented in international conferences (e.g., ICIS, ECIS, MCIS, and EURO).

1 Introduction

The proliferation of online travel websites (e.g., TripAdvisor, Yelp) facilitates the creation and dissemination of user generated content (UGC) and plays a significant role in the decision making of the tourists. Previous researches recognise UGC of online travel websites as a crucial resource and various studies have used them to acquire knowledge including: the examination of the relationship of UGC with revenue and reputation (Luca, 2016), the management of online image (O'Connor, 2010), how UGC build trust (Jeacle and Carter, 2011), the influential role of UGC in tourists' decision making process (Filieri and McLeay, 2014), the measurement of hospitality satisfaction (Limberger et al., 2014; Calero-Sanz et al., 2022), tourism analytics based on the comments (Miah et al., 2017), etc. Therefore, UCG in the tourist industry fuel research studies from different scientific disciplines ranging from marketing and decision making to information systems and strategy.

An in-depth analysis on UGC raised questions regarding the validity of the studies in terms of the data reliability, process and outcomes. Orlikowski and Scott (2014) studied from a sociomateriality perspective how valuation changes when conducted online and suggested that the existing literature usually employs objective measures that are not sufficient to portray such a complex sociotechnical phenomenon. They proposed that valuation is constituted in practice and impacts both visitors (online travel websites) and employees in hospitality industry (e.g., restaurant, hotel, airline). Ranking entrepreneurship is a relevant concept recommended by Rindova et al. (2018) who suggested the creation of a new research stream, where specific actors will propose ranking mechanisms, in order to reveal the required details and better understand the impacts of valuation in general. Indeed, Orlikowski and Scott (2014) pointed out that the lack of details regarding TripAdvisor's valuation mechanism (namely TripAdvisor popularity ranking), is an important obstacle to thorough study and suggested that in many respects such ranking doesn't correspond well to literature expectations.

Ganzaroli et al. (2017) indicated that, TripAdvisor contributes to the popularity of high-quality restaurants yet, they point out that TripAdvisor's algorithm which is designed to reward quality, doesn't always achieve that. TripAdvisor (2020) briefly describes that its ranking mechanism materialises three factors, namely quality, recency and quantity. In regards to the first factor, it establishes the quality of the experience through the rating system (named bubbles in TripAdvisor's terms). For the second factor, a weighting scheme is incorporated in order to prioritise latest comments as more significant and valid. Finally, the last factor is utilised to understand the confidence of the ranking based on the amount of the available comments that are associated with a specific hospitality area. However, the specific algorithm is not published.

All these suggest that there is a gap in ranking mechanisms either due to the lack of information on the details of the ranking algorithms or the use of objective measures (stars/bubbles) which cannot capture the details provided by UGC (Kim et al., 2019). The objective of this work is to suggest a ranking mechanism for organisations in tourism industry (e.g., hotels, restaurants) which transforms UGC information resources into topics and views every topic as performance criteria. This ranking mechanism combines topic modelling, to extract the evaluation criteria (topics), and multi-criteria decision making (MCDM) to compare the performance and generate the rankings of the involved business entities. A commonly used method for extracting topics is latent Dirichlet allocation (LDA) which is applied in this paper. It is a semi-supervised approach, because a few parameters, either at the side of topic modelling (i.e., the number of topics) or at the stage of MCDM (e.g., the importance of positive vs. negative comments), are not known in advance and human intervention is required to drive the method.

The proposed ranking algorithm was implemented on a sample of restaurants located in Athens, Greece. Restaurants have not received significant research attention, compared to hotels, and according to Ong (2012), the findings are less accurate than hotels' due to the subjective and variable nature of restaurant comments. To this end, restaurants are an attractive, yet challenging, context for study. Athens, Greece was selected as a country where the primary industry is tourism and a major factor that contributes to its wealth is its restaurants. Also, the availability of an industry expert to review the results of this research motivated us to select the target market in Athens Greece. The information resources (e.g., comments, ratings) were collected from TripAdvisor.

In this study, it was observed a couple of issues regarding ratings and comments. First, we define *rating divergence* the misalignment between the comments (text) and the corresponding rate (numbers/stars). Star rating influences the users' decision-making process (Wan and Nakayama, 2022) yet, the occasional rating divergence (e.g., typing error, strict/generous rating), is very instructive because it signals to avoid a ranking mechanism design exclusively based on ratings. It should be clarified that in case of TripAdvisor, when users are provided suggestions on the best restaurants according to their preferences (food, location, etc.) they assume that these derive from the ratings that a restaurant has received from previous users. This is not the case as the ranking offered by TripAdvisor actually derives from a ranking algorithm which is not known in detail. In our study, we observed a few cases where tourists had a negative dining experience, in high ranked restaurants. Through their comments we noticed a specific pattern: the negative commentary seems to primarily reflect the level of disappointment from the deceiving ranking obtained through TripAdvisor and secondarily the disappointment from the restaurant itself. We call this case *deceiving ranking*, to define a bad dining experience rating which sometimes encompasses the disappointment from the deceiving ranking offered through a platform (i.e., TripAdvisor).

The rest of the paper is organised as follows. First, we provide a review of contemporary studies in online restaurant reviews, briefly present the topic modelling field and a review of existing multi-criteria techniques. Next, we present the research methodology of this study and propose a ranking mechanism. Within this new ranking scheme, we performed variations to examine how they fit to the existing ratings provided by the tourists. Finally, we draw conclusions and posit future research streams.

2 Literature review

Information is a main resource in tourist industry and has significantly altered the way it operates. Most platforms (e.g., TripAdvisor, Yelp), during the preparation phase of a travel, facilitate searching and booking and encourage, through crowdsourcing or recommendation practices, experience sharing after the service provision (e.g., Yelp). Managing information resources is crucial for all stakeholders in travel industry and we limit the scope of information resources only on the UGC delivered by the tourists in regards to restaurants. However, even this delimited scope includes highly dimensional and unstructured data and a way to address the challenges is by using topic modelling techniques. Finally, ranking the performance of restaurants is possible through the employment of MCDM. In the following subsections are thoroughly presented: UGC for restaurants, topic modelling and MCDM.

2.1 Online restaurant reviews

Internet has penetrated many households and it is the means for a lot of people to seek experience-based information on products and services they need (Yang, 2017) from the comfort of a preferred location and a choice of time. UGC facilitates awareness and decision making on desired products and services (Nilashi et al., 2018) and is a considered a valuable resource. In regards to choosing which restaurant to visit, this option has become very attractive since there is an abundance of positive or negative comments, ratings and information related to them. Restaurant UGC influence potential customers to visit a restaurant or not (Parikh et al., 2014; Filieri et al., 2021), therefore, the value of opinions expressed in UGC is indisputable. However, the immense available information, in order to provide value needs to be processed and filtered so it is easy to understand and accessed. For instance, the desired result from hundreds of comments on restaurants is for a person to identify the best restaurant that will cover his/ her needs and for a business to identify what customers like and dislike. To this point, UGC have received some criticism especially for their fairness and integrity. Comment viewers are concerned about businesses masquerading as independent commenters to post 'fake' entries or they are annoyed by comments which do not correspond to realistic assessments (Pezenka and Weismayer, 2020; Wu et al., 2020; Krishnan and Wan, 2021). Schuckert et al. (2015) found significant differences on the rating behaviour between language groups of English and non-English speaking users. On the same token, Nakayama and Wan (2018) argue that people's ethnic origin and their culture guide their sentiment expressions which are in turn reflected upon their comments. In other words, although UGC are an important tool for businesses and customers it is evident that both comments and their respective star ratings sometimes confuse the readers and they do not always reflect the truth.

It appears that the purpose of restaurant related UGC researches focus either on customer satisfaction or emotions. Indicatively, studies have analysed restaurant related UGC, to underpin restaurant satisfiers (Bilgihan et al., 2018); to emerge the best method analysing reviews (Laksono et al., 2019; Line et al., 2020); to identify the influence or review attributes and sentiments on star ratings (Gan et al. 2017); to extract emotions and frequently mentioned dining aspects (Luo and Xu, 2019); to identify the factors affecting the perceived usefulness of online reviews (Liu and Park, 2015); to uncover and compare the satisfaction of tourists in restaurants from their restaurant ratings and reviews (Jia, 2019); to identify influencing factors on restaurant customers' revisit intention (Yan et al., 2015); and to propose a new model for text analysis to improve inference and prediction of customer ratings (Büschken and Allenby, 2016).

Content analysis is a research method of analysing written and verbal messages with respect to understand the nuance and semantics of words, sentences and documents (Miles et al., 2014) and such textual collection is usually noted as corpus. Krippendorff (2018) identified the characteristics of contemporary content analysis, opposed to traditional content analysis efforts originated 60 years back, and argued that it is a research methodology beyond of counting qualitative data. The intrinsic characteristic of analysing data sources both from quantitative and qualitative perspective has made content analysis a fundamental approach when dealing with text. One of the most central tasks in content analysis is text categorisation, which is done either by hand (manually) or computer-aided. In the former case, the researcher follows a top-down or a bottom-up approach to code data (Urquhart, 2012), while the latter utilises computing to systematically analyse content. For example, Lei and Law (2015) followed computer-aided content analysis (using NVivo) to study 615 comments from 22 different restaurants at Macau. In this work, the coding of the corpus was manually performed by one researcher and provided accurate and insightful results. However, the traditional content analysis is a resource consuming approach with limits regarding the volume of the comments reviewed.

A favourable research approach to analyse UGC in tourism industries has been sentiment analysis (Fuentes-Moraleda et al., 2020; Tang et al., 2019; Jain et al., 2021). Extensive studies on natural language processing (NLP) have proposed tools and techniques that enable the analysis of large amounts of data and categorise text based on positive, negative or neutral attitude. There are three different levels (also called units of analysis) of sentiment analysis. In the document level, it is examined if a comment expresses a positive, negative or neutral sentiment (Pang et al., 2002; Valdivia et al., 2018). In the sentence level, it is examined if each sentence contains a positive, negative or neutral opinion (Wiebe et al., 1999; Valdivia et al., 2018). In word level, using lexicon-based approaches and by counting the word's polarity, a comment is classified as positive or negative (Cambria et al., 2017). It is common to formulate sentiment analysis as a binary or multi-class problem depending on the level of analysis selected by the researcher. However, there are cases where within a document different aspects are discussed and for some of them there is a positive opinion while for others a negative one. In this case, classifying the comment as positive or negative would cause to miss the valuable information encapsulated within the comment. On this token, at the entity and aspect level the focus is to recognise all the sentiment expressions within a document as well as which aspects they refer to (Feldman, 2013). In other words, within the scope of sentiment analysis it is possible that a sentence might have a positive sentiment about an object and at the same time a negative one on another object (Liu, 2012); therefore, it is ambiguous to determine how a consumer perceived a service provision. It is common that sentiment analysis is joint with a classical classification method (e.g., naïve Bayes, support vector machines) in order to automatically classify a comment (at document level) based on the lexical and/or non-lexical text classification of the sentiments (Kirilenko et al., 2018).

Factor analysis is a collection of techniques used to study the relationships between variables (observed or latent). When dealing with real data the number of variables is immense and there is a need to reduce the number of dimensions. Topic modelling addresses the issue of reducing the dimensions especially when the data are in the form of a corpus (Titov and McDonald, 2008). Based on the aforementioned technique, indicatively Liu and Park (2015) examined the factors affecting the perceived usefulness of online customer reviews. Whereas Büschken and Allenby (2016) proposed a new model for text analysis based on the sentence structure to improve inference and prediction of consumer ratings.

2.2 *Topic modelling*

Topic modelling is considered an unsupervised technique addressing a significant problem in important research fields such as NLP, information retrieval and content analysis. The task of Topic modelling is the extraction of topics (represented as a set of words) that occur in a collection of documents. In more detail, a document is a sequence of N words $w = (w_1, w_2, \ldots, w_N)$while the collection of M different documents shape a corpus $D = (w_1, w_2, \ldots, w_M)$. The lowest unit of analysis is a word and as a discrete element it is considered to be part of a vocabulary vector V.

LDA (Blei et al., 2003) generalises the LSI through the utilisation of a probabilistic model of categorising documents into topics. In this approach, the underlying assumption is that the data arise from a generative probabilistic process that includes hidden variables, while at a next step it is proposed a hidden structure using posteriori inference (e.g., to evaluate the conditional distribution of the hidden variable over the observations). LDA utilises probability distributions to model the associations:

1 documents with topics

2 topics with words.

In the remainder of this section, we present in more detail the LDA approach since it was the technique that was utilised in our research.

LDA presumes that a document might include multiple topics. Blei et al. (2003) argue that there is an imaginary generative process and based on that the authors create a document d. Each document is associated with different (latent) topics (noted as t), using words w from a finite and fixed vocabulary *V*. Each topic is defined from a discrete distribution of words and different topics have different mixtures of words, but they are all selected from the same *V*. Finally, LDA assumes the existence of k different topics in the corpus *D*.

Taking into account the aforementioned approach, LDA uses the following notion for the involved random variables. The mixture proportions for the topics is noted as β_k, where each value of the vector is a distribution over the vocabulary *V*. The total topic proportions for a document d is denoted as θ_d and subsequently $\theta_{d,k}$ is the proportion of the k^{th} topic for document d. LDA uses the hidden random variable z_d to model (indices) the assignments of topics over a document and in more detail the $z_{d,n}$ shows the topic assignment of the nth word of the d document. In addition, the observed nth word of a document d is noted as $w_{d,n}$ and it is the only observed random variable that the user supplies. Finally, LDA employs as hyper-parameters:

1 α is the Dirichlet prior represented as a k-vector to approximate θ_d

2 β is also a Dirichlet prior to approximate the topics β_k.

Blei and Lafferty (2006) suggest the LDA probabilistic generative process as Figure 1 depicts.

Figure 1 LDA generative process

LDA Algorithm
1. For each *topic*, draw a distribution over words $\beta_k \sim Dir(\beta)$
2. For each *document*,
a. Draw a vector of *topic* proportions $\theta_d \sim Dir(a)$
b. For each *word*,
i. Draw a *topic* assignment $z_d \sim Mult(\theta_d),\ z_{d,n} \in \{1, \dots, K\}$
ii. Draw a *word* $w_{d,n} \sim Mult(\beta_{z_{d,n}}),\ w_{d,n} \in \{1, \dots, V\}$

To this end, the efficient analysis of a corpus requires the examination of the posterior distribution of the topics (β_k), topic proportions (θ_d) and topic assignments z_d conditioned on the document *d*.

Topic modelling with LDA (or variations) is a popular option when dealing with reviews from online platforms related to tourism. For instance, Guo et al. (2017) used LDA to identify key dimensions of customer service as expressed by hotel visitors. Taecharungroj and Mathayomchan (2019) examined the tourists' reviews in TripAdvisor for Phuket's attractions and Brand et al. (2017) adopted LDA to support business analytics and gain insights regarding smart urban tourism. Therefore, it is evident that LDA is a powerful technique that has already been employed in the tourist sector.

The present study contributes to the aforementioned literature in the following manner:

a it studies reviews from restaurants which contain a more challenging vocabulary (set of words) to manage, taking into account that gastronomy includes a wide variety of low frequency-high value words (e.g., the special plates offered by a restaurant, local plates)

b the topics are utilised in a multi-criteria analysis and as a result it is required to deliver topics that make sense and are also supported by proper weighting scheme.

The next section provides details regarding the latter issue.

2.3 Multi-criteria decision making

Decisions by optimising a given task, based on unique criteria, to maximise benefits or minimise costs are vital within complex organisational environments (Gandibleux, 2006). However, this very complexity does not allow decision-makers to provide an optimal solution based on individual criteria, due to conflicting objectives within an organisation. Therefore, multiple criteria have to be considered which are often contradictory or incommensurable. Specifically, the essence of multi-criteria optimisation is the selection of the best from a set of alternatives (Gandibleux, 2006).

There are two main approaches for the analysis of multi-criteria problems. The value system approach which is based on a quantitative methodology and the non-compensatory approach which builds relations that tolerate the incomparability among decision actions (outranking relation) (Siskos and Spyridakos, 1999).

MCDM models aid the evaluation of the overall preference values of alternatives with respect to multiple criteria (Choo et al., 1999). Based on the approach of the evaluation, the models are distinguished in four categories depending on whether they evaluate alternative priorities, outrank the alternatives, optimise of the distance or perform a mix of the previous ones (Pohekar and Ramachandran, 2004).

Pardalos et al. (1995) proposed an alternative grouping for multi-criteria approaches as follows: multi-objective linear programming (Steuer, 1986), multi-attribute utility theory (Keeney and Raiffa, 1993), outranking relations theory (Roy, 1968) and preference disaggregation approach (Jacquet-Lagreze and Siskos, 1982, 2001). To this end, it is evident that there are no distinct approaches in regards to multi-criteria as the field is still expended and researched upon. Additionally, Ho et al. (2010) mentioned cases where the fuzzy set theory (introduced by Zadech, 1965) is applied in multi-objective linear programming and multi-attribute utility theory. The pair $(X, \mu_X(x))$, $x \in X$ is called a fuzzy set, where $\mu_X(x)$ is the membership function of X and takes values between 0 and 1 (Zadech, 1965). When T is an ordinary subset of X, the pair $(T, \overline{T})$ is a partition of X provided that $T \neq \emptyset$ and $T \neq X$. When T is a fuzzy set ($\neq \emptyset$, $\neq X$) the pair $(T, \overline{T})$ is called fuzzy partition; more generally an m-tuple $(T_1, T_2, \ldots, T_m)$ of fuzzy sets ($\forall i$, $T_i \neq \emptyset$ and $T_i \neq X$) such that:

$$\forall x \in X, \sum_{i=1}^{m} \mu_x(T_i) = 1 \text{ (orthogonality)}$$

is still called a fuzzy partition of X [Dubois, (1980), p.13].

In this study, we use topics (T_i) as fuzzy criteria whose membership function $\mu_x(T_i)$ counts the level of participation of a topic in the users' comments of a restaurant (x). The objective is to classify the restaurants based on the users' total experience (X) as it is expressed through their comments.

Over different approaches for investigating comments, this research utilised LDA methodology which produces topics for the analysis of the data. These topics consist of words where, words with high appearance probability dictate what each topic is about. In every language, a word has multiple meanings and that causes a degree of fuzziness which is more intense when considering a sequence of words. For instance, the word 'price' in conjunction with the words 'low, medium, high' differentiates the original word adding a scale. However, since the word price is the base which is combined with three other possible words, it would appear more often than the others. Thus, the word 'price' will be at the top of the probabilities of a topic whereas the three others may not appear.

In another instance, words from each comment maybe located in all or some topics which means that each topic may, in some rate, participate or not in a comment. Therefore, there is fuzziness within each comment. Thus, it is evident that since there are no crisp values we resort to fuzzy values. In other words, we use fuzzy logic which encompasses membership function to obtain the ranking of the restaurants.

3 Methodology

This study proposes a method that:

1 in-depth analyses restaurant comments

2 suggest ranking restaurants using multiple criteria based on the comments.

The former is achieved through the employment of cross-industry standard process for data mining (CRISP-DM) (Shearer, 2000) and the latter by adjusting well established ranking techniques suggested within the MCDM such as ELECTRE and Promethee (e.g., Sevkli, 2010; Behzadian et al., 2010; Rouyendegh and Erol, 2012; Corrente et al., 2013; Botti and Peypoch, 2013; Botti et al., 2020). In the next paragraphs, we briefly present the research method and the proposed ranking scheme employed in this work.

3.1 Research method

This study was facilitated by a unified framework involving tasks associated with both data mining and multi-criteria decision. On the one hand, the objective of CRISP-DM was to analyse content and reveal dimensions (also called as topics or aspects) associated with user ratings, and on the other hand, the objective of MCDM was to propose a ranking-scheme that directly compares the restaurants based on topics and provide a more accurate view of the market. CRISP-DM is an iterative process model that organises a data-mining project around six phases and it was employed to facilitate this study.

Initially, we observed discrepancies between users' comments and restaurant ratings. This motivated us to suggest a rating scheme that considers the content of the comments alongside the rating and the pricing level of a restaurant.

At a next phase (data understanding), we collected from tripAdvisor.com data regarding the top 10% restaurants (based on search results from TripAdvisor) located in Athens, Greece. The collected data were restaurant comments (e.g., date of review, comment title, comment body) and details regarding the restaurant (e.g., pricing, cuisine specialisation).

In regards to the data preparation phase of CRISP-DM, we performed few tasks (manual or automated) such as correcting the misspelled words, removing stop words and punctuation, tokenisation of the corpus, etc. During data preparation, we decided to split the entire sample into two disjoint groups based on the users' ratings. As a result, we derived a dataset with positive restaurant review comments (those that have rating >= 4) and a second dataset with the remaining comments (rating <= 3) that disclosed one or more negative aspects of a restaurant. Such decision was aligned with the objective of this work which was to review in detail all positive and negative reviews and extract topics associated with the rate of the comment. During this data preparation phase, we set sampling parameters in order to avoid under fitting and over fitting of the data mining model (i.e., we set an upper limit for popular/high-commented restaurants).

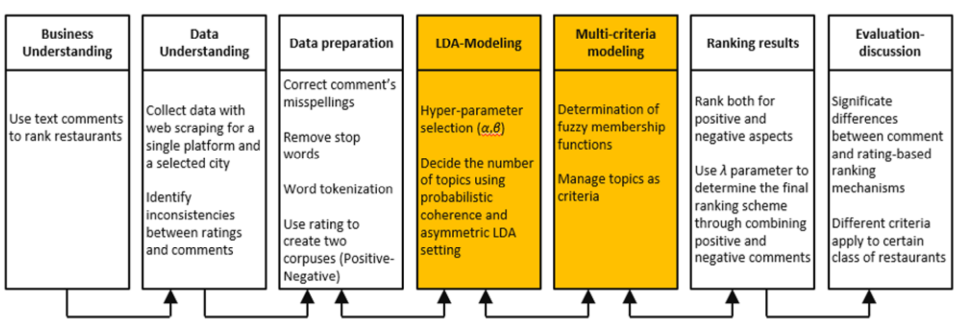
The next step of CRISP-DM is the modelling phase and we used LDA to extract topics. Specific tasks regarding the capitalisation of LDA include the fine tune of the hyper-parameters, examination of n-grams models, the evaluation of topics' coherence measure, etc. The results of the LDA were then supplied to MCDM since the evaluation phase was performed in collaboration with MCDM.

In the step of multi-criteria modelling, the fuzzy data extracted from the LDA, which was a table with the values of topics' membership function, was processed to create superiority relations. In this process, the higher values from the membership function in a topic rise to the top since they are superior to the lower values. This takes place for each topic separately but also for all the topics of an alternative solution (each restaurant) together. This way a classification emerged.

The evaluation of the positive and the negative superiority relations created respectively positive (descending order) and negative (ascending order) ranking results. Finally, from the process of the previous two derived the final ranking results. This qualitative type ranking is used in contrast to the quantitative ranking (benchmark) which is based on the average of the stars provided by the users (benchmark star rating).

Figure 2 summarises the research method we followed. The white boxes represent tasks within the context of CRISP-DM and the orange boxes are tasks that correspond to MCDM.

Figure 2 Research method (see online version for colours)



The data collection was facilitated by a web crawler designed for that purpose. It was implemented in R using rvest (Wickham, 2022) and RSelenium (Harrison, 2022) packages in a two stages process. Firstly, a search for restaurants at Athens, Greece was performed and based on that list, the comments associated with each restaurant were recursively downloaded. RSelenium as a browsing automation tool was essential for the creation of events to extract the whole length of the comment. The data collection was conducted at the end of February 2020 and it took about three days to complete.

3.2 Proposed ranking mechanism

In this section, it is briefly described the proposed ranking mechanism (scheme). Beforehand, it should be noted that there were some assumptions made for this research. It was assumed that: all comments are valid and of equal importance; there is a strong relevance between the comment and the rate; the number of topics in unknown beforehand; the topics are adequate to formulate criteria and the topics are considered fuzzy variables.

Initially, we separated the available corpus D into two parts where, positive comments shaped the D^+ corpus and the negative comments were included within the D^- corpus. The underlying idea of splitting corpus D into a positive and a negative part, was to gain a detailed view on how tourist experience their restaurant visits and the extracted topics were considered either positive or negative. Keeping a single corpus D is a suitable approach when dealing with document classification or sentiment analysis; however, in this work, we needed an in-depth investigation of positive and negative reviews, and consequently we separated topics to positive (T_i^+) or negative (T_i^-) experiences taking into account what the respective content described.

An important LDA result is the θ_d that describes the distribution of topics over a document. Very low levels (e.g., smaller than 0.005) of θ_d were ignored and the θ_d was normalised for each document such that the sum of probabilities for a document d was equal to 1. We viewed restaurants as a collection of documents and we evaluated the average value θ_d for all the documents related to each different restaurant. The above process was repeated twice, separately for the positive and the negative comments.

In this research, subsets (T_1, T_2, …, T_m) are fuzzy variables (topics) that represent TripAdvisor users' experiences that have visited proposed restaurants. Additionally, $\mu_x(T_i)$ is the membership function of these topics in a restaurant. The $\mu_x(T_i)$ is defined as the average probability of the appearance of a topic in the comments of a restaurant. The vector of this $\mu_x(T_i)$ are the rows of $c_{n,i}^{\{+,-\}}$ matrix. The sum of the probabilities of topics' appearance equals to 1 for each restaurant (orthogonality). Thus, all topics are a partition from the total TripAdvisor users' restaurant experience. Finally, X defines the total experience of all users from all restaurants they have rated and it is a fuzzy set.

In an ideal situation, in regards to multi criteria analysis, the optimum solution would be the one that dominates over others in all criteria. Unfortunately, this is not feasible in most cases since, comparing alternative solutions, it is evident that one is better than another on certain criteria. Thus, since is not feasible to obtain an ideal solution, we seek a feasible solution. However, this requires a measure that will countify how one alternative excels another. In our case, we define a measure that will countify the superiority of each criterion as well as cumulatively among two alternatives. Y and Z are two fuzzy numbers where $Y = \mu_Y(T_1)/T_1 + \dots + \mu_Y(T_m)/T_m$, $Z = \mu_Z(T_1)/T_1 + \dots + \mu_Z(T_m)/T_m$. Y is superior to Z in T_i if $\mu_Y(T_i) \geq \mu_Z(T_i)$ and Y is superior to Z in X if $\mu_Y(T_i) \geq \mu_Z(T_i)$, $\forall i = 1, \dots, m$ where fuzzy subsets (T_1, T_2, …, T_m) ($\forall i$, $T_i \neq \emptyset$ and $T_i \neq X$) are a partition of X.

Table 1 Symbols and definitions

Symbol	*Definition*
N, n	The set of documents, the n^{th} restaurant
$D^{\{+,-\}}$	Corpus containing
$d_{n,j}$	The j^{th} document (comment) for the n^{th} restaurant
$K^{\{+,-\}}$	The dimension of topics (positive or negative)
$T_i^{\{+,-\}}$	The set of topics derived either from positive or negative documents (comments) with $k \in \{1 \ldots K\}$
θ_d	The assignment (probability) of document over topics based on the LDA
$c_{n,i}^{\{+,-\}}$	The average assignment of n^{th} restaurants over i^{th} topic either positive or negative
$\mu_x(T_i)$	The topics' membership functions as given by LDA
$w_i^{[+,-]}$	The weights of topics $w_i \in c_d^{[*]}$, *=+, – as given by LDA either positive or negative
s	Superiority function among alternatives
S_p	Total superiority of alternative p against all the rest
R_p	The ranking function of alternative p
λ	The ratio between the negative and positive ranking functions
Y, Z	Fuzzy numbers that define the users' total experience for the restaurants y and z.

If (w_1, w_2, …, w_m) are weight sets of (T_1, T_2, …, T_m) respectively, where $\sum_{i=1}^{m} w_i = 1$, then we define the superiority of Y over Z as the real number:

$$s = \sum_{i \in I} w_i, \text{ where } I = \{1 \le i \le m : Y \text{ are superior of } Z \text{ in } T_i \text{ in } X\}$$

In other words, we define criteria weights to calculate cumulative superiority. This idea derives from ELECTRE methods. In this research case, the criteria are represented by topics and we deemed necessary to use as weights the normalised occurrence of topics.

Finally, as it is not adequate to just obtain the superior alternative, we define a cumulative measure that will rank the obtained alternatives.

Let the set of fuzzy numbers: $Y_j = {}^{\mu_{Y_j}(T_1)}\!/\!_{T_1} + \cdots + {}^{\mu_{Y_j}(T_m)}\!/\!_{T_m}$, $j = 1, \ldots, n$, then the total superiority of Y_p to Y_j, $i \neq p$ is

$$S_p = \sum_{i \neq p} s_i, \text{ where } s_i \text{ is the superiority of } Y_p \text{ to } Y_i,\ i = 1, \ldots, n$$

It should be highlighted that we perform two different rankings, one for the topics with positive comments and another for the topics with negative comments. Then, we obtain a final ranking function subtracting positive and negative scores.

$$R_p = (1-\lambda) * S_p^{+} - \lambda * S_p^{-},\ p = 1, \ldots, n,\ \lambda \in (0, 1)$$

To this end, the proposed ranking algorithm, as Figure 3 presents, combines LDA (steps 1–4) and the multi-criteria analysis with fuzzy numbers (steps 5–9).

Summarising, in our proposed method, emphasis is given on the magnitude of the participation of a comment in the topics. The higher presence of a comment in the negative topics lowers the restaurant's ranking and contrary the higher presence of a comment in the positive topics gives them a higher rating.

Figure 3 Ranking restaurants based using comments

Proposed ranking algorithm
1. $\{D^+, D^-\} \leftarrow$ Prepare and split Corpus based on the rating
2. $\beta_i^+ \leftarrow$ LDA topics with positive documents $\beta_i^- \leftarrow$ LDA topics with negative documents
3. $\theta_d^+ \leftarrow$ Normalize assignments for the positively related documents $\theta_d^- \leftarrow$ Normalize assignments for the negatively related documents
4. For all documents $d_{n,j}$ of restaurant *n* Calculate the average assignment of restaurants over positive topics $c_{n,i}^+ = \sum_{j=1}^{J} \frac{\theta_d^+}{J}, \forall\, \theta_d^+ > 0$ Calculate the average assignment of restaurants over negative topics $c_{n,i}^- = \sum_{j=1}^{J} \frac{\theta_d^-}{J}, \forall\, \theta_d^- > 0$
5. s$\leftarrow$ pairwise grading of alternatives
6. $S_p^+ \leftarrow$ calculate the positive ranking function for p alternative
7. $S_p^- \leftarrow$ calculate the negative ranking function for p alternative
8. $R_p \leftarrow$ calculate the total ranking function for p alternative based on the ratio (λ) of the negative over the total comments
9. Sorting using R_p

4 Experimental results

4.1 Data availability and setting

The context of this study is restaurants in Athens, Greece and tripadvisor.com has been the platform where the comments were collected from. The data collection was enabled by a web crawler and it was conducted the last week of June 2020. Starting from the main restaurants' section of TripAdvisor, we set as target the accumulation of a 10% sample from all restaurants. We collected:

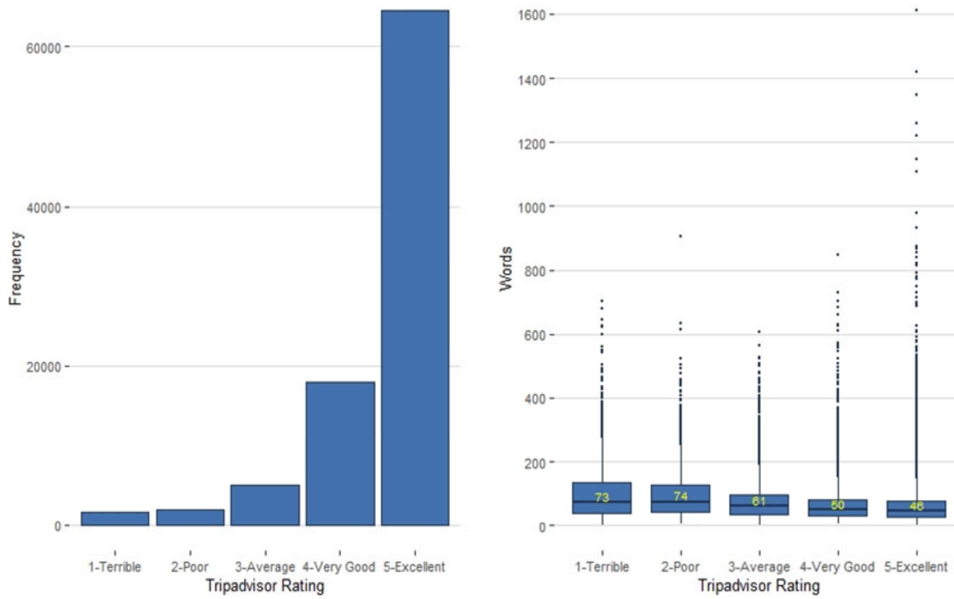
1 restaurant details (e.g., name, pricing, cuisine details)

2 the associated comments (e.g., review title, text comment, rating).

An initial corpus (*D*) of 300 restaurants (out of 2,270 restaurants in the area) and 91.085 comments was formed.

Some interesting descriptive statistics of *D* include the distribution of ratings for the sampled comments and the number of words (comment length). In more detail, the comments were negative skewed and most rated a very good or an excellent experience according to the tripadvisor's scale, ranging from 1 (terrible) to 5 (excellent) [see Figure 4(a)]. Our results confirm Jamshidi et al. (2019), who observed that the assigned ratings are frequently more positive than they should be. Regarding the number of words, we observed that a bad experience generates a longer comment on average (e.g., the median for a Terrible rating is 73 words) while very extensive comments (more than 800 words) that are considered outliers (and perhaps with suspicious motives) are found at excellent ratings [see Figure 4(b)]. In general, the relation between the ratings and the comments' length is a U-shaped distribution.

Figure 4 Rating distributions (a) frequency and (b) words used (see online version for colours)



(a) (b)

As other relevant studies suggest (Jia, 2019; Luo and Xu, 2019) the unit of analysis is a user's single comment (named as document in the topic modelling community) and it is the grain to create the *D* corpus. However, the overwhelming positive comments can hinder/limit the formation of the negative aspects during the topic modelling task and in order to address such issue, we split *D* into a positive D^+ and a negative $D^–$ corpus. Using comments' ratings, we decided to include into the positive corpus (D^+) all documents (comments) with a rate above or equal to 4 and the remaining documents populated the negative corpus ($D^–$). Further, we adopted some heuristic rules to filter whether a comment will (or not) participate in the corpus. The filters can be classified into two broad categories:

1 restaurant performance

2 comments' length.

The former is based on the observation that the number of comments varied significantly between restaurants so we set an upper and a lower limit threshold. Restaurants that did not meet the lower limit were excluded from the corpus and at the high end were considered restaurants with 350 comments at most. The latter is based on the observation that lengthy comments (> 800 words) have outlier characteristics and were therefore, excluded. The filters were mostly applied during the formation of the positive corpus (D^+).

Text operations were identical to both corpuses and included: lowering the letters of the words, removal of stop words, punctuations and numbers, creation of a custom dictionary with insignificant words (i.e., Athens, restaurant), elimination of rare words of the corpus, etc. Table 2 summarises some key characteristics of the two corpuses. In both, the number of restaurants sampled is close to 12% of the total available restaurants. The comments (documents) included in D^+ is significant higher due to the skewed distribution of rating. Finally, the unique words (vocabulary) refer to the number of words that topic modelling methods utilised, and the increased number found in D^- is the result of the lengthier and more detailed comments when a negative experience is reported.

Table 2 Characteristics of the two corpuses employed

	Positive experience corpus (D^+)	*Negative experience corpus (D^-)*
Relevant restaurants	277	279
Number of comments (documents)	51,967	8,351
Number of words (vocabulary)	9,011	4,371

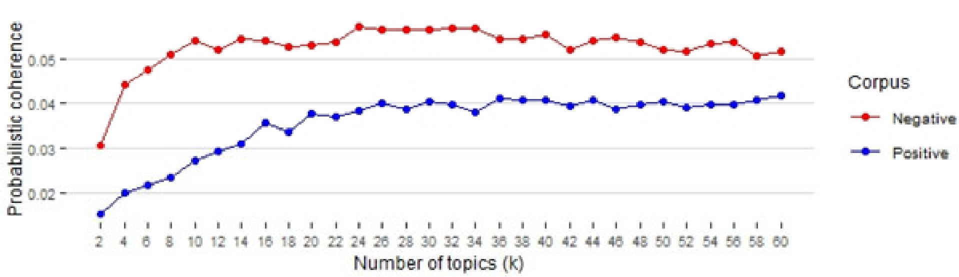
4.2 LDA setting

This section addresses the creation and comparison of different LDA models and issues regarding the LDA setting, namely the selection of the number of topics and the hyper-parameter (α, β) setting.

An important LDA parameter is the number of topics (k), which should be decided based on the specific problem and by utilising prior knowledge. In our study, we did not have enough information to set an appropriate level for k and we decided to create multiple LDA models for a different number of topics. Therefore, we created a series of LDA models starting from the minimum number, which is $k = 2$, up to a sufficient number, which was set as $k = 60$. We set the initial values of hyper-parameters aligned to Griffiths and Steyvers (2004) propositions.

What makes a good topic is difficult to pinpoint precisely and various metrics have been proposed to address the issue, such as perplexity (Blei et al., 2003) and probabilistic coherence metrics [e.g., UCI by Newman et al. (2010) and UMass by Mimno et al. (2011)]. Perplexity, in our dataset, was found to be unstable and we turn our attention towards coherence metrics, where we found probabilistic topic modelling as the most suitable. Averaging all available topic coherences, it was possible to have a single measure at the LDA model level, which corresponded to the k parameter. The suggestion is to select the k^{th} model that maximises the probabilistic topic coherence. Figure 5 illustrates the aforementioned process and each dot on the lines represents an individual LDA model, trained at the corresponding corpus.

Figure 5 Selecting number of topics using probabilistic coherence (see online version for colours)



The results suggested a low level of probabilistic coherence. Taking into account the underlying calculation method, two major causes related to the corpus vocabulary could be drawn. Firstly, the vocabulary associated with each corpus is considerable large, thus a significant number of words have low probability (frequency). Secondly, a significant part of the words are highly correlated (e.g., diner-night) but they also statistical independent due to the semantics. High correlation might be falsely caused when a word is part of a restaurant's name (e.g., Acropolis tavern) and/or the proportion of comments among restaurants. To this end, we found that for the positive corpus (D^+) an adequate selection was 36 topics and regarding the negative corpus (D^-) 24 topics.

Table 3 Indicative topics and the α Dirichlet prior

Corpus	*a*	*Topic*	*Top five terms*	*Topic interpretation*
Positive corpus (k = 36)	0.218	Topic 1	Found, reviews, tripadvisor, decided, meal	The decision has been based on existing positive reviews in TripAdvisor, which is very common within such platforms
	0.152	Topic 4	Wine, service, experience, dining, excellent	A diner with nice wine, among others, offered a very nice experience.
	0.027	Topic 33	Service, price, quality, excellent, top	A valuable restaurant selection because it combines different dinning aspects
Negative corpus (k = 24)	0.241	Topic 12	Service, wait, time, table, minutes	Long waiting time for a table. Such issue is very frequent especially during the summer holidays.
	0.132	Topic 17	Reviews, tripadvisor, average, service, disappointed	Tripadvisor's reviews mislead the expectations and the tourists express disappointment
	0.033	Topic 11	Hard, card, rock, pay, credit	Problems with the payment system of the restaurant

For setting LDA we followed the propositions of Wallach et al. (2009). On the one hand, the α Dirichlet prior was calculated using an optimisation scheme during the Gibbs sampling iterations and on the other hand we set $\beta = 0.05$ for all the words. Profoundly, the α Dirichlet prior express the researcher's belief about how a topic is distributed among the document and the higher the value the more probable is a document to contain a topic. Indicatively, Table 3, illustrates three topics per corpus, the asymmetric value of a, the related topic followed by the top five most frequent terms and a brief explanation by an industry expert. The expert noted that the derived topics were intriguing based on his experience in the subject matter. A skepticism was expressed in regards to more specialised topics such as one of them that related to Indian restaurants which was not expected. Also, according to the expert, some topics were overlapping but this occurred due to the meaning expressed by the sequence of the words and the probabilistic assignment used by the method. All in all, the expert found the results of the process interesting and in line with customers' perceptions.

In Table 3, the topics were selected based on asymmetric α Dirichlet prior different levels of values (highest, median and minimum). For example regarding the positive corpus, Topic 1 has the highest α value, topic 4 is close to the median, etc. The asymmetric value of α also found to have a linear relationship with the topic prevalence measure, so at a generic level it is possible to argue that a high α value indicates how much of a comment is associated with a topic. The values of α in the negative corpus are lower on average compared to positive corpus. Through reviewing the data, we concluded that the negative corpus has qualitative difference compared to the positive corpus and in more detail the former is more precise and factual, while the positive corpus tends to express overall experience and perceptions with generic words. It should also be noted that topic 1 in the positive corpus reveals that the readers value highly TripAdvisor's (in this case positive) reviews but the same also stands for negative reviews which was evident in our topic 17 in the negative corpus.

4.3 Ranking all restaurants

In this section, we propose a multi-criteria mechanism that ranks restaurants based on the comments and the ratings provided by the users. Before delving into the details, we shortly discuss the transformation of the LDA results to a manageable format from the multi-criteria method, as it is the interface between two research fields. An important random variable of LDA is the per-document topic proportions (θ_d). Considering that a collection of documents refers to a restaurant, it is possible to group LDA's assignments at a restaurant level, through the capitalisation of a simple descriptive function (e.g., median) for the assignment's distribution. Figure 6 exhibits such transformation and shows the topics assignment for 14 anonymised restaurants for both corpuses.

Figure 6 Average assignment probability of topics over restaurants (rest. and top. refers to restaurant and topic accordingly) (see online version for colours)

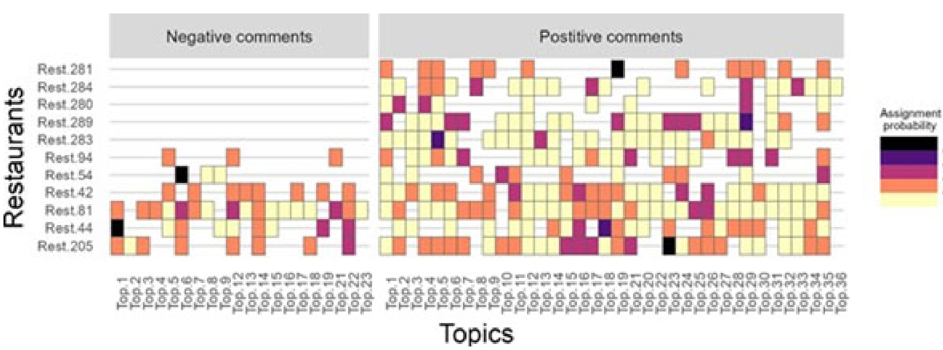


Figure 6 reviewed either by columns or rows, provides some initial findings. Firstly, reading by columns it is possible to identify which topics have been substantially discussed (colour tiles of the heat map) and which topics have been lightly discussed or not at all (no colour) per restaurant. Very low levels (e.g., smaller than 0.005) of θ_d were ignored. Additionally, the more heavily/more frequently discussed topics are noted with darker colour tiles. Secondly, reading by rows it is possible to identify how the topics are discussed in each restaurant. Similar to the first case, the topics which have been substantially discussed are noted with coloured tiles whereas, those which have been lightly discussed or not at all are the ones which do not have any colour. The more heavily/more frequently discussed topics are noted with darker colour tiles. The θ_d was normalised for each document such that the sum of probabilities for a document d was equal to 1 (orthogonality). Either examining Figure 6 by rows or columns, what is more important is the intensity of a topic (darker colour tiles) because it signifies a high average assignment probability of topics over restaurants and indicates either high quality service (for positive comments) or areas that need immediate improvement (for negative comments).

Similarly, viewing the heatmap by rows, an overview is acquired regarding the mixture of positive-negative aspects per different restaurant. For example, Rest.44 has a mixture of both positive and negative reviews and it seems having a significant problem regarding topic 1 (in negative corpus) but it also excels in topic 18. In the contrary, Rest.281 has not any negative assignments and it seems that most of the positive comments are discussed in topic 19. By managing topics as criteria, it is feasible to switch from the topic modelling area to MCDM and consequently create an efficient ranking mechanism.

Table 4 Negative topics weights

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	…	T_{24}
w	0.04	0.02	0.07	0.04	0.02	0.09	0.02	0.05	0.02	0.02	…	0.05

Table 5 Average assignment of the n^{th} restaurant over the i^{th} topic $c_{n,i}^{\{+,-\}}$ (negative comments)

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	…	T_{24}
Rest. 1	0.01	0.01	0.04	0.04	0.03	0.15	0.06	0.08	0.01	0.01	…	0.02
Rest. 2	0.07	0.02	0.06	0.03	0.01	0.12	0.04	0.01	0.03	0.03	…	0.02

Assuming that each topic is a different criterion, the LDA exports the matrix θ_d from which derive the tables $c_{n,i}^{\{+,-\}}$ containing the average probability of the appearance of a topic in the comments of a restaurant. The average probability is the membership function $\mu_{T_i}(x)$ of each topic. For each restaurant, the membership functions of its topics occupy a row in matrix $c_{n,i}^{\{+,-\}}$. However, some multi-criteria methods require use of the criteria weights as well. In our study, the weights are the normalised values of the topics prevalence $w_i \in c_d^{[*]}, * = +, -$ and the subsets (T_1, T_2, …, T_m) are fuzzy variables (topics). As an example let us consider the negative comments for two restaurants to distinguish superiority of the one over the other.

Table 4 depicts the weights of each topic. Table 5 contains the values of negative topics' membership functions for restaurants 1 and 2 (Rest.1, 2). The columns of this matrix are the topics' membership functions. In other words, the intersection of the 3rd row and 9th column is the value of the membership function for the 9th topic in the 2nd restaurant and equals to 0.03. For example, we calculate where Rest.1 is superior to Rest.2 in Table 5 as follows:

$$s = \sum_{i \in I} w_i, \text{ where } I = \left\{1 \le i \le 24 : \text{Rest. 1 is superior to Rest.2 in } i^{\text{th}} \text{ topic}\right\} = 0.51$$

$$I = \{4, 5, 6, 7, 8, 11, 13, 15, 16, 17, 18, 20, 22, 24\}$$

Similarly, we proceed with calculation on all pairs of alternatives to create relevant *s*. Then, all these s for each restaurant are summed based on function:

$$S_n^{[-]} = \sum_{i \neq n} s_i, \text{ where } s_i \text{ is the superiority of } n_{th} \text{ restaurant over all the rest}$$

This way we have the negative ranking. Relatively, we work to calculate the positive ranking.

Finally, we subtract the obtained values of the negative from the positive rankings, over λ, to obtain the final ranking:

$$R_n = (1-\lambda) * S_n^{+} - \lambda * S_n^{-}$$

where λ is the rate that we want negative comments to be considered over the positive ones. If λ is equal to 0.5 the positive and negative comments are distributed evenly.

Table 6 Partial ranking results of our method for the top rankings for each λ and some of the last in the ranking

Restaurant	*Star ranking mechanism*			*Proposed ranking mechanism*				
	Low stars (1 2 3)	*High stars (4 5)*	*Average stars ranking*	*λ = 0.1*	*λ = 0.2*	*λ = τ0.3*	*λ = 0.4*	*λ = 0.5*
Rest.42	18	363	43	1	1	12	24	44
Rest.228	76	2,089	44	2	2	14	29	50
Rest.11	43	261	229	3	5	31	59	101
Rest.81	67	448	208	4	11	32	54	92
Rest.94	2	49	51	23	3	3	9	15
Rest.54	1	39	39	34	4	4	10	16
Rest.281	0	33	1	146	45	1	1	1
Rest.282	0	68	3	157	63	2	2	2
Rest.284	0	42	4	182	97	5	3	3
Rest.280	0	60	10	185	101	8	4	4
Rest.289	0	30	5	195	120	9	5	5

Table 6 depicts how rating changes based on λ using our data. We calculated multiple λ, where for λ equal to 0.1, 10% weight of the negative comments over the positive ones is considered, etc., and included the average star ranking of the corresponding restaurant from TripAdvisor as well as the total stars per restaurant in two groups: low stars (1 to 3 stars) and high stars (4 to 5 stars). The results indicate that a restaurant ranks higher than the average star ranking for $\lambda = 0.1$ (e.g., 42, 228, 11, 81).

However, an issue arises whether it is better to consider isobaric the negative and positive comments (λ equal to 0.5, 0.4) or, since the negative comments are considerably less, to take the negative comments proportionally to the positive ones (λ equal to 0.1, 0.2, 0.3).

It is evident from Figure 6 and Table 6 that for $\lambda = 0.5$ restaurants with minimal or not at all negative comments precede others. As observed from Table 6, there is a divergence of the proposed mechanism over the benchmark one, when $\lambda = 0.1$, $\lambda = 0.2$. In other words, there is a divergence when the negative comments are considered proportionally to the positive ones whereas going towards isobaric consideration of negative and positive comments the results from both mechanisms converge.

5 Discussion

In the big data era, the mass collected information becomes meaningful through proper processing while technology affects users' ratings (Orea-Giner et al., 2022). Such a processing is ranking mechanisms which are a new research area and it is evident that there are issues which need to be addressed (Rindova et al., 2018). It is not known how current ranking mechanisms work exactly since they are not published and therefore it is not known if they are efficient (Orlikowski and Scott, 2014). This gap is very stressful for management since the lack of evidence implies inability to make the right decisions. In the case of restaurants there are even less researches due to the complexity that arise from the subjective and variable nature of the users' comments. In this study, the comments of restaurants' users in TripAdvisor were the resources used for multi criteria decision making.

TripAdvisor's popularity ranking mechanism materialises quality (rates), recency (latest ratings) and quantity of comments. The proposed methodology adopts recency as an important factor that identifies properly the current situation. However, there were issues that emerged considering quality and quantity of comments. In regards to quality, we encountered an occasional divergence among the rating provided by a user and its corresponding comment. In regards to quantity, it was observed that restaurants with a large number of comments (some negative among many positive) were ranked, according to TripAdvisor, higher than a restaurants that had overall fewer comments but they were all positive. In other words, excellent restaurants that do not have a significant mass of comments are undermined in the shuffle. Furthermore, since lesser quality restaurants are suggested higher, more customers visit them and consequently more comments are generated. The exponential distribution of comments resembles the Matthew effect (Rigney, 2010) where a mass amount of comments leads to further advantage (additional comments) and restaurants with a relative small amount of comments strive to overcome the commentary shortage. As a result, a vicious cycle is created that cancels the significance of quantity. Brought together quality and quantity mishaps in TripAdvisor's ranking mechanism result in a deceiving ranking and users highlight that, when they comment that the proposed ranking does not correspond to reality. On the other hand, it should be noted that other users praise TripAdvisor's ranking. Actually, both negative and positive dispositions towards TripAdvisor appear equally in this study's topics and reveal that there are some issues that derive partially from the stars' rating and more from comments' volume.

In the proposed ranking mechanism, any volume of comments is acceptable as long as the sample is statistically significant and other than that the actual volume does not affect the ranking. Instead, all negative and positive comments are analysed to extract the ranking of the restaurants. Through the research it became evident that λ is a factor that determines the ranking. Specifically, the issue is that users tend to complain but still give 3 or more stars which in turn classifies the comment as positive. When λ is given a lot of weight (considering λ equal to 0.4 and 0.5) the ranking we derive through the comments is closer to the average star ranking. Contrariwise, when the negative comments are considered proportionally to the positive ones (λ equal to 0.1, 0.2, 0.3) the results deviate from the average star ranking. This rises a dilemma whether or not to consider the negative comments isobaric to the positive ones or take them proportionally. Since, when taking negative and positive comments isobarically converges towards stars ratings this study proposes to weight positive and negative comments equally and regardless of the comments' volume. In other words, we propose a ranking based on the λ which is between 0.4 and 0.5 and we deviate from considering more significant the large volume of the positive comments but we handle them as if they have equal weight with the fewer negative comments. Otherwise, the negative comments are not surfaced which leads to comments such as "selection of place was due to very high ratings from TripAdvisor …the place is awful ...I recommend to avoid this place". After all, it is the negative comments that contribute to improvements. This is in line with Mehraliyev et al. (2020) who claim that the negative experiences had higher effect on customer rating.

The proposed ranking mechanism is derived solely from the users' comments and significantly differs from TripAdvisor's popularity ranking. In the former, comments were handled at a document level with a qualitatively manner in order to produce topics which are used as performance criteria. This method expands the dominant trend where the overall restaurant rating or the perceived value of the dining experience is associated with limited value aspects such as food, atmosphere and service (Gao and Xu, 2022) which is the case with TripAdvisor. Another difference is related to the participation of the rating scale. In our case, the rates were only used to split the corpus into positive and negative comments. Contrary, TripAdvisor's popularity ranking claims that rating along with recency and comments' quantity are the important factors that constitute their own ranking approach.

The utilisation of LDA in the context of tourist industry was found valuable because it transforms unstructured data (comments) into a valuable resource that supplies the ranking mechanism. In our case, the higher probabilistic coherence of the negative corpus (D^-) indicates that complaints tend to be more precise and factual (e.g., "we waited 30 minutes for a table"), compared to the positive experiences where the use of generic words describing perceptions might exist across different topics (e.g., "we had a lovely night"). Therefore, we suggest topics extracted from the negative comments as a source of pitfalls to avoid and corrective actions to undertake whereas, the positive topics as a confirmatory basis of what tourists are looking for. Regarding LDA's results, the high yielded number of topics urges a thorough study that will identify and remove insignificant topics for the ranking process.

The proposed ranking mechanism differs significantly from the one implemented by TripAdvisor as the former utilises qualitative data (comments) whereas the later utilises quantitative data (bubbles, volume of comments and recency). In particular, the results cannot be compared to TripAdvisor's because firstly, it does not exist a common sample of test data to use and secondly, the TripAdvisor's algorithm is not published in detail so that it can be replicated. Therefore, we can only compare our results with the average of the stars. The results suggest that there are serious discrepancies. Furthermore, direct comparison with other studies is not possible due to the use of different datasets as well as a different focus. Comparisons could be partially made only with the LDA results (comparing extracted topics).

6 Conclusions

We have proposed a ranking mechanism based on the qualitative characteristics of tourists' comments. We have argued that the qualitative aspects of comments are more instructive compared to the quantitative scale rating. The proposed mechanism is based on LDA to create the evaluation dimensional space, and the ranking follows the principles of MCDM by utilising a λ parameter. We examined the proposed model for sample restaurants using the TripAdvisor platform and evaluated the results compared to a simple quantitative star rating scheme.

6.1 Theoretical implications

From the theoretical perspective, our study is a step forward in ranking methods. It makes four main contributions toward formulating a more accurate and fair ranking system.

First, it is not affected by the volume of the comments which is proven to yield false results, as is the case with TripAdvisor, but it considers the qualitative characteristics of them. Additionally, the proposed ranking method weights positive and negative comments equally regardless of their volume. This ensures that the fewer negative comments are not undermined. Second, the proposed ranking mechanism does not consider rates (stars/ bubbles) which are one dimensional and do not correspond to the written comments. They are only utilised to split comments into positive and negative. Third contribution is the utilisation of LDA to transform comments into meaningful data that supply the ranking mechanism. Finally, it introduces a new ranking algorithm based on fuzzy logic using fuzzy variables to handle the complexity of the topics that derive from LDA.

6.2 Practical implications

From the practical perspective, the proposed improved ranking mechanism pertains to recommendation agent platforms, the users of such platforms and business management.

A more accurate ranking mechanism contributes to the credibility of a recommendation agent platform, increasing its popularity over other similar platforms. The proposed ranking mechanism also benefits restaurant management as each business gets the rank it deserves consolidating those who offer better value for money while alerting the ones who get a lower ranking to investigate possible issues that need to be addressed. Aside the first two, the study benefits mostly the users by eliminating possible deceptions while searching for a particular aspect that they are interested in. A better ranking mechanism where results can be trusted contributes to making faster and better-informed decisions.

All in all, the proposed ranking mechanism simplifies the decision-making process and contributes to the market competitiveness.

6.3 Limitations and future lines for research

We acknowledge that this work has some limitations. As earlier discussed, we consider the proposed mechanism as preliminary, thus we recognise the incorporation of a sliding time window that weighs the most recent comments as more significant/accurate, which is an important characteristic to enhance the validity of the results. Moreover, the validation of the study is based on a benchmark mechanism which assumes that most of the platforms examine the quantitative (star rating) rather than the qualitative (comments) aspect of ranking, as this work suggests. The adopted benchmark rating scheme suggests the lack of a ranking ground truth employed by the platforms, therefore we consider the generalisation of the findings as limited. Lastly, it should be noted that the volume of positive comments is considerably higher than that of negative comments which may influence the data balance.

We suggest that future works will deal with alternative ranking mechanisms utilising the qualitative aspects of user-generated content. Regarding the future improvements of the proposed method, the options are few: study and evaluate n-gram language models, examine fuzzy topic modelling, apply the method for hotels, attractions, etc. Additionally, the large number of topics yielded by LDA urges a study that will identify and remove insignificant topics.

References

Behzadian, M., Kazemzadeh, R.B., Albadvi, A. amd Aghdasi, M. (2010) 'PROMETHEE: a comprehensive literature review on methodologies and applications', *European Journal of Operational Research*, Vol. 200, No. 1, pp.198–215.

Bilgihan, A., Seo, S. and Choi, J. (2018) 'Identifying restaurant satisfiers and dissatisfiers: Suggestions from online reviews', *Journal of Hospitality Marketing & Management*, Vol. 27, No. 5, pp.601–625.

Blei, D. and Lafferty, J. (2006) 'Correlated topic models', *Advances in Neural Information Processing Systems*, Vol. 18, p.147.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) 'Latent Dirichlet allocation', *Journal of Machine Learning Research*, January, Vol. 3, pp.993–1022.

Botti, L. and Peypoch, N. (2013) 'Multi-criteria ELECTRE method and destination competitiveness', *Tourism Management Perspectives*, Vol. 6, pp.108–113.

Botti, L., Petit, S. and Zhang, L. (2020) 'Strategic decision concerning tourist origins portfolio: a decision process based on the ELECTRE method and applied to French Polynesia', *Tourism Economics*, Vol. 26, No. 5, pp.830–843.

Brandt, T., Bendler, J. and Neumann, D. (2017) 'Social media analytics and value creation in urban smart tourism ecosystems', *Information & Management*, Vol. 54, No. 6, pp.703–713.

Büschken, J. and Allenby, G.M. (2016) 'Sentence-based text analysis for customer reviews', *Marketing Science*, Vol. 35, No. 6, pp.953–975.

Calero-Sanz, J., Orea-Giner, A., Villacé-Molinero, T., Muñoz-Mazón, A. and Fuentes-Moraleda, L. (2022) 'Predicting a new hotel rating system by analysing UGC content from tripadvisor: machine learning application to analyse service robots influence', *Procedia Computer Science*, Vol. 200, pp.1078–1083.

Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. (2017) 'Affective computing and sentiment analysis', in *A Practical Guide to Sentiment Analysis*, pp.1–10, Springer, Cham.

Choo, E.U., Schoner, B. and Wedley, W.C. (1999) 'Interpretation of criteria weights in multicriteria decision making', *Computers & Industrial Engineering*, Vol. 37, No. 3, pp.527–541.

Corrente, S., Greco, S. and Słowiński, R. (2013) 'Multiple criteria hierarchy process with ELECTRE and PROMETHEE', *Omega*, Vol. 41, No. 5, pp.820–846.

Dubois, D.J. (1980) *Fuzzy Sets and Systems: Theory and Applications*, Vol. 144, Academic Press.

Feldman, R. (2013) 'Techniques and applications for sentiment analysis', *Communications of the ACM*, Vol. 56, No. 4, pp.82–89.

Filieri, R. and McLeay, F. (2014) 'E-WOM and accommodation: an analysis of the factors that influence travelers' adoption of information from online reviews', *Journal of Travel Research*, Vol. 53, No. 1, pp.44–57.

Filieri, R., Yen, D.A. and Yu, Q. (2021) '# ILoveLondon: an exploration of the declaration of love towards a destination on Instagram', *Tourism Management*, Vol. 85, p.104291.

Fuentes-Moraleda, L., Diaz-Perez, P., Orea-Giner, A., Munoz-Mazon, A. and Villace-Molinero, T. (2020) 'Interaction between hotel service robots and humans: a hotel-specific service robot acceptance model (sRAM)', *Tourism Management Perspectives*, Vol. 36, p.100751.

Gan, Q., Ferns, B.H., Yu, Y. and Jin, L. (2017) 'A text mining and multidimensional sentiment analysis of online restaurant reviews', *Journal of Quality Assurance in Hospitality & Tourism*, Vol. 18, No. 4, pp.465–492.

Gandibleux, X. (Ed.) (2006) *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*, Vol. 52, Springer Science & Business Media.

Ganzaroli, A., De Noni, I. and van Baalen, P. (2017) 'Vicious advice: analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice', *Tourism Management*, Vol. 61, pp.501–510.

Gao, Y. and Xu, A. (2022) 'How do aspects of chain restaurants affect the overall rating: TripAdvisor multi-dimensional rating system analysis', *Journal of International Technology and Information Management*, Vol. 31, No. 1, pp.97–124.

Griffiths, T.L. and Steyvers, M. (2004) 'Finding scientific topics', *Proceedings of the National academy of Sciences*, Vol. 101, No. Suppl. 1, pp.5228–5235.

Guo, Y., Barnes, S.J. and Jia, Q. (2017) 'Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation', *Tourism Management*, Vol. 59, pp.467–483.

Harrison, J. (2022) *RSelenium: R Bindings for 'Selenium WebDriver'. R Package Version 1.7.7* [online] http://docs.ropensci.org/RSelenium (accessed 30 July 2022).

Ho, W., Xu, X. and Dey, P.K. (2010) 'Multi-criteria decision making approaches for supplier evaluation and selection: a literature review', *European Journal of Operational Research*, Vol. 202, No. 1, pp.16–24.

Jacquet-Lagreze, E. and Siskos, J. (1982) 'Assessing a set of additive utility functions for multicriteria decision-making, the UTA method', *European Journal of Operational Research*, Vol. 10, No. 2, pp.151–164.

Jacquet-Lagreze, E. and Siskos, Y. (2001) 'Preference disaggregation: 20 years of MCDA experience', *European Journal of Operational Research*, Vol. 130, No. 2, pp.233–245.

Jain, P.K., Pamula, R. and Srivastava, G. (2021) 'A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews', *Computer Science Review*, Vol. 41, p.100413.

Jamshidi, S., Rejaie, R. and Li, J. (2019) 'Characterizing the dynamics and evolution of incentivized online reviews on Amazon', *Social Network Analysis and Mining*, Vol. 9, No. 1, p.22.

Jeacle, I. and Carter, C. (2011) 'In TripAdvisor we trust: rankings, calculative regimes and abstract systems', *Accounting, Organizations and Society*, Vol. 36, Nos. 4–5, pp.293–309.

Jia, S. (2019) 'Measuring tourists' meal experience by mining online user generated content about restaurants', *Scandinavian Journal of Hospitality and Tourism*, Vol. 19, Nos. 4–5, pp.371–389.

Keeney, R.L. and Raiffa, H. (1993) *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Cambridge University Press, Cambridge.

Kim, H., Joun, H.J., Choe, Y. and Schroeder, A. (2019) 'How can a destination better manage its offering to visitors? Observing visitor experiences via online reviews,' *Sustainability*, Vol. 11, No. 17, p.4660.

Kirilenko, A.P., Stepchenkova, S.O., Kim, H. and Li, X. (2018) 'Automated sentiment analysis in tourism: comparison of approaches', *Journal of Travel Research*, Vol. 57, No. 8, pp.1012–1025.

Krippendorff, K. (2018) *Content Analysis: An Introduction to its Methodology*, 4th ed., SAGE Publications Inc., Thousand Oaks.

Krishnan, K. and Wan, Y. (2021) 'The detection of fake reviews in bestselling books: exploration and findings', *Journal of Electronic Commerce in Organizations (JECO)*, Vol. 19, No. 4, pp.64–79.

Laksono, R.A., Sungkono, K.R., Sarno, R. and Wahyuni, C.S. (2019) 'Sentiment analysis of restaurant customer reviews on TripAdvisor using naïve Bayes', in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, IEEE, July, pp.49–54.

Lei, S. and Law, R. (2015) 'Content analysis of TripAdvisor reviews on restaurants: a case study of Macau', *Journal of Tourism*, Vol. 16, No. 1.

Limberger, P.F., Dos Anjos, F.A., de Souza Meira, J.V. and dos Anjos, S.J.G. (2014) 'Satisfaction in hospitality on TripAdvisor. com: an analysis of the correlation between evaluation criteria and overall satisfaction', *Tourism & Management Studies*, Vol. 10, No. 1, pp.59–65.

Line, N.D., Hanks, L. and Dogru, T. (2020) 'A reconsideration of the EWOM construct in restaurant research: what are we measuring?', *International Journal of Contemporary Hospitality Management*.

Liu, B. (2012) 'Sentiment analysis and opinion mining', *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, pp.1–167.

Liu, Z. and Park, S. (2015) 'What makes a useful online review? Implication for travel product websites', *Tourism Management*, Vol. 47, pp.140–151.

Luca, M. (2016) *Reviews, Reputation, and Revenue: The case of Yelp.com*, Com, 15 March, Harvard Business School NOM Unit Working Paper, (12-016) (accessed 8 August 2020).

Luo, Y. and Xu, X. (2019) 'Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: a case study of yelp', *Sustainability*, Vol. 11, No. 19, p.5254.

Mehraliyev, F., Kirilenko, A.P. and Choi, Y. (2020) 'From measurement scale to sentiment scale: examining the effect of sensory experiences on online review rating behavior', *Tourism Management*, Vol. 79, p.104096.

Miah, S.J., Vu, H.Q., Gammack, J. and McGrath, M. (2017) 'A big data analytics method for tourist behaviour analysis', *Information & Management*, Vol. 54, No. 6, pp.771–785.

Miles, M.B., Huberman, A.M. and Saldana, J. (2014) *Qualitative Data Analysis: A Methods Sourcebook*, SAGE Publications.

Mimno, D., Wallach, H., Talley, E., Leenders, M. and McCallum, A. (2011) 'Optimizing semantic coherence in topic models', in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, July, pp.262–272.

Nakayama, M. and Wan, Y. (2018) 'Is culture of origin associated with more expressions? An analysis of Yelp reviews on Japanese restaurants', *Tourism Management*, Vol. 66, pp.329–338.

Newman, D., Noh, Y., Talley, E., Karimi, S. and Baldwin, T. (2010) 'Evaluating topic models for digital libraries', in *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, pp.215–224.

Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E. and Alizadeh, A. (2018) 'Travelers decision making using online review in social network sites: a case on TripAdvisor', *Journal of Computational Science*, Vol. 28, pp.168–179.

O'Connor, P. (2010) 'Managing a hotel's image on TripAdvisor', *Journal of Hospitality Marketing & Management*, Vol. 19, No. 7, pp.754–772.

Ong, B.S. (2012) 'The perceived influence of user reviews in the hospitality industry', *Journal of Hospitality Marketing & Management*, Vol. 21, No. 5, pp.463–485.

Orea-Giner, A., Fuentes-Moraleda, L., Villacé-Molinero, T., Munoz-Mazón, A. and Calero-Sanz, J. (2022) 'Does the implementation of robots in hotels influence the overall TripAdvisor rating? A text mining analysis from the Industry 5.0 approach', *Tourism Management*, Vol. 93, p.104586.

Orlikowski, W.J. and Scott, S.V. (2014) 'What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector', *Organization Science*, Vol. 25, No. 3, pp.868–891.

Pang, B., Lee, L. and Vaithyanathan, S. (2002) *Thumbs Up? Sentiment Classification using Machine Learning Techniques*, arXiv preprint cs/0205070.

Pardalos, P.M., Siskos, Y. and Zopounidis, C. (1995) *Advances in Multicriteria Analysis*, Kluwer Academic Publishers, Dordrecht.

Parikh, A., Behnke, C., Vorvoreanu, M., Almanza, B. and Nelson, D. (2014) 'Motives for reading and articulating user-generated restaurant reviews on Yelp.com', *Journal of Hospitality and Tourism Technology*.

Pezenka, I. and Weismayer, C. (2020) 'Which factors influence locals' and visitors' overall restaurant evaluations?', *International Journal of Contemporary Hospitality Management*.

Pohekar, S.D. and Ramachandran, M. (2004) 'Application of multi-criteria decision making to sustainable energy planning – a review', *Renewable and Sustainable Energy Reviews*, Vol. 8, No. 4, pp.365–381.

Rigney, D. (2010) *The Matthew Effect: How Advantage Begets further Advantage*, Columbia University Press.

Rindova, V.P., Martins, L.L., Srinivas, S.B. and Chandler, D. (2018) 'The good, the bad, and the ugly of organizational rankings: a multidisciplinary review of the literature and directions for future research', *Journal of Management*, Vol. 44, No. 6, pp.2175–2208.

Rouyendegh, B.D. and Erol, S. (2012) 'Selecting the best project using the fuzzy ELECTRE method', *Mathematical Problems in Engineering*.

Roy, B. (1968) 'Classement et choix en présence de points de vue multiples', *Revue française d'informatique et de recherche opérationnelle*, Vol. 2, No. 8, pp.57–75.

Schuckert, M., Liu, X. and Law, R. (2015) 'A segmentation of online reviews by language groups: how English and non-English speakers rate hotels differently', *International Journal of Hospitality Management*, Vol. 48, pp.143–149.

Sevkli, M. (2010) 'An application of the fuzzy ELECTRE method for supplier selection', *International Journal of Production Research*, Vol. 48, No. 12, pp.3393–3405.

Shearer, C. (2000) 'The CRISP-DM model: the new blueprint for data mining', *Journal of Data Warehousing*, Vol. 5, No. 4, pp.13–22.

Siskos, Y. and Spyridakos, A. (1999) 'Intelligent multicriteria decision support: overview and perspectives', *European Journal of Operational Research*, Vol. 113, No. 2, pp.236–246.

Steuer, R.E. (1986) *Multiple Criteria Optimization: Theory, Computation and Applications*, John Wiley and Sons, New York.

Taecharungroj, V. and Mathayomchan, B. (2019) 'Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand', *Tourism Management*, Vol. 75, pp.550–568.

Tang, F., Fu, L., Yao, B. and Xu, W. (2019) 'Aspect based fine-grained sentiment analysis for online reviews', *Information Sciences*, Vol. 488, pp.190–204.

Titov, I. and McDonald, R. (2008) 'Modeling online reviews with multi-grain topic models', in *Proceedings of the 17th International Conference on World Wide Web*, April, pp.111–120.

TripAdvisor (2020) *Everything You Need to Know about the Tripadvisor Popularity Ranking [web page]*, 15 August [online] https://www.tripadvisor.com/TripAdvisorInsights/w765.

Urquhart, C. (2012) *Grounded Theory for Qualitative Research: A Practical Guide*, Sage, Thousand Oaks, CA.

Valdivia, A., Luzón, M.V., Cambria, E. and Herrera, F. (2018) 'Consensus vote models for detecting and filtering neutrality in sentiment analysis', *Information Fusion*, Vol. 44, pp.126–135.

Wallach, H.M., Mimno, D.M. and McCallum, A. (2009) 'Rethinking LDA: why priors matter', in *Advances in Neural Information Processing Systems*, pp.1973–1981.

Wan, Y. and Nakayama, M. (2022) 'A sentiment analysis of star-rating: a cross-cultural perspective', in *Proceedings of the 55th Hawaii International Conference on System Sciences*, January.

Wickham, H. (2022) *rvest: Easily Harvest (Scrape) Web Pages* [online] https://rvest.tidyverse.org/, https://github.com/tidyverse/rvest (accessed 30 July 2022).

Wiebe, J., Bruce, R. and O'Hara, T.P. (1999) 'Development and use of a gold-standard data set for subjectivity classifications', in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, June, pp.246–253.

Wu, Y., Ngai, E.W., Wu, P. and Wu, C. (2020) 'Fake online reviews: literature review, synthesis, and directions for future research', *Decision Support Systems*, Vol. 132, p.113280.

Yan, X., Wang, J. and Chau, M. (2015) 'Customer revisit intention to restaurants: evidence from online reviews', *Information Systems Frontiers*, Vol. 17, No. 3, pp.645–657.

Yang, F.X. (2017) 'Effects of restaurant satisfaction and knowledge sharing motivation on eWOM intentions: the moderating role of technology acceptance factors', *Journal of Hospitality & Tourism Research*, Vol. 41, No. 1, pp.93–127.

Zadech, L. (1965) 'Fuzzy sets', *Information and Control*, Vol. 8, No. 3, pp.338–353.