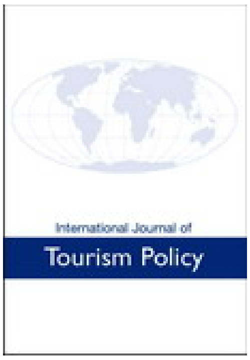


International Journal of Tourism Policy

ISSN online: 1750-4104 - ISSN print: 1750-4090
https://www.inderscience.com/ijtp

Aspect-based sentiment analysis: Jamie's Italian restaurant case study

Joana Figueira, Bráulio Alturas, Ricardo Ribeiro

DOI: 10.1504/IJTP.2023.10056703

Article History:

Received:	14 October 2021
Last revised:	30 June 2022
Accepted:	30 June 2022
Published online:	13 July 2023

Aspect-based sentiment analysis: Jamie's Italian restaurant case study

Joana Figueira and Bráulio Alturas*

Instituto Universitário de Lisboa (ISCTE-IUL),
ISTAR-Iscte (University Institute of Lisbon),
Av. Forças Armadas, 1649-026, Lisboa, Portugal
ORCID: 0000-0003-0142-3737
ORCID: 0000-0002-2058-693X
Email: joana_figueira@iscte-iul.pt
Email: braulio.alturas@iscte-iul.pt
*Corresponding author

Ricardo Ribeiro

Instituto Universitário de Lisboa (ISCTE-IUL),
INESC-ID Lisboa,
Av. Forças Armadas, 1649-026, Lisboa, Portugal
Email: ricardo.ribeiro@iscte-iul.pt

Abstract: Consumers use technologies to share their experiences, leading to the creation of online platforms where the main objective is to allow users to share their opinion about products or services, such as hotels, books, restaurants, and search for the opinions of other users. The emergence of these online platforms has changed the business dynamics, the restaurant sector was no exception. The main goal of this work is to understand how different factors impact the final review rating of a restaurant, using two Jamie Oliver restaurants as a case study. A model was applied that allows us to identify the such factors and their associated sentiment through text mining methods. Using this model, it was possible to understand which factors influence the rating the most. Results show that the factors most mentioned in the reviews were 'food' and 'service' and the least mentioned were 'atmosphere' and 'location'.

Keywords: online reviews; text mining; restaurants; sentiment analysis; Jamie Oliver.

Reference to this paper should be made as follows: Figueira, J., Alturas, B. and Ribeiro, R. (2023) 'Aspect-based sentiment analysis: Jamie's Italian restaurant case study', *Int. J. Tourism Policy*, Vol. 13, No. 4, pp.315–330.

Biographical notes: Joana Figueira holds a BSc in Industrial Management and Logistics and a MSc in Information Systems Management both from Instituto Universitário de Lisboa (ISCTE-IUL) (University Institute of Lisbon), Lisbon. Her scientific research interests encompass logistics, operations management, supply chain management and information systems management. She is currently working as Process Lead at Philip Morris International.

Bráulio Alturas (PhD) is an Associate Professor of Department of Information Science and Technology of Instituto Universitário de Lisboa (ISCTE-IUL) (University Institute of Lisbon), Lisbon, and researcher of the Information

Systems Group of ISTAR (Information Sciences, Technologies and Architecture Research Center). He holds a PhD in Management with specialisation in Marketing, a MSc in Management Information Systems, and a BSc in Business Organization and Management, all from ISCTE-IUL. His scientific research interests and publications are in acceptance and use of technology, digital marketing, social media, e-commerce, information management, information systems, direct selling and consumer behaviour.

Ricardo Ribeiro is an Associate Professor of Department of Information Science and Technology of Instituto Universitário de Lisboa (ISCTE-IUL) (University Institute of Lisbon), Lisbon, where he is the coordinator of the Artificial Intelligence scientific area, and an integrated researcher at INESC-ID Lisboa, working on Human Language Technologies. His current research interests focus on high-level information extraction from unrestricted text, speech or music, and improving machine-learning techniques using domain-related information. He has participated in several European and Nationally-funded projects and scientific events, either as organiser or as member of the program committee (IJCAI, ICASSP, LREC, Interspeech). He was the editor of a book on the computational processing of Portuguese.

1 Introduction

The Web 2.0 revolution included a change in consumers' behaviour, as such nowadays consumers have an 'active participation' in content creation. Moreover, it also changed the way companies interact with consumers (Constantinides and Fountain, 2008). One of the major consumers' behaviour changes was the rise of 'experience sharing' about services and/or products social media, which led to the appearance of websites and online applications whose objective is to allow users to write their own reviews about products or services, such as hotels, books, and restaurants, and to search for other users' opinions (Park and Gretzel, 2007).

The emergence of these online platforms changed the business sector dynamics, and the restaurant sector is no exception. On these types of platforms, users can share their experiences, photos, and classify restaurants. This type of platform is great for consumers, but also for restaurants, since they allow restaurants to share the general ambiance, their food menu, and prices. Having good reviews is a good way to attract new consumers (Dubey, 2016).

The research question proposed for this study is the following: 'Which factors most influence the rating of a restaurant from online reviews?' There are several studies about reviews, but their focus is essentially on the impact of published opinion for future customers and not the benefits and impacts for the supplier. One of the sectors where this topic has been the most discussed is the hotel sector. Related to this problem statement, this study has the main objective of performing sentiment analysis, using text mining tools, to understand how each of the different factors mentioned by the users influences the final rating of the review. This work focused on Zomato and Tripadvisor reviews.

Consumer attitudes and behaviours are fundamentally dynamic processes; therefore, understanding consumer dynamics is crucial to truly understanding consumers' behaviour and for firms to formulate appropriate actions. Recent history in empirical marketing research has enjoyed increasingly richer consumer data because of technology and the conscious data collection efforts of firms. Richer data, in turn, have propelled the development and application of quantitative methods in modelling consumers' dynamics and have contributed to the understanding of complex dynamic behaviours across many domains (Zhang and Chang, 2021). Hence, the importance of studies like this in studying consumers' behaviour from the data that consumers themselves produce.

The specific objectives of this study are the following: to establish the dimensions/categories – factors – of the reviews analysis; to identify the main topics mentioned in the reviews; to apply a sentiment analysis model for the mentioned topics; and, to create a performance matrix.

The case study chosen uses data about two Jamie Oliver restaurants, located in two different countries – Portugal and England. These two restaurants were chosen since the one in England is closed while the other in Portugal is still open, being able to conclude which are strengths and weaknesses of both restaurants and understand which factors contribute the most to customer satisfaction.

The Lisbon restaurant opened in January 2018 and continues to operate. It is a 500m2, divided between 3 floors, which can accommodate 174 people in seats in the 3 rooms and 2 terraces. There is also a private room where you can host small private events.

The Soho London restaurant ran very successfully from 2008 to 2019. In May 2019, the Jamie Oliver Restaurant Group went into administration. As a result, 22 of Oliver's 25 remaining Jamie's Italian restaurants have closed, and 1000 people lost their jobs. It was a huge loss for the celebrity chef, who at one point had 42 Jamie's Italian restaurants in the UK.

2 Literature review and conceptual background

2.1 Online platforms

Technological advances in communication have caused significant changes in how companies communicate with customers, but also in how customers communicate with each other (Parreira et al., 2021). Online platforms allow consumers to share their opinion related to companies, products, or services. Companies saw, in this environment, also a good way to reach users and potential consumers by being present, either through personal profiles or through fan pages (Ferreira and Alturas, 2010). The need to aggregate the different types of businesses or sectors led to the development of several websites and mobile apps, with the main objective of bringing together companies and consumers in one single space. The advantage for companies is that they can search the contents that concern them, being able to analyse the satisfaction and which offers satisfy the needs of the consumers. Another advantage for companies is that by adopting these communication channels, they have practically free advertising, they can transmit quality, guarantee the most experienced consumers trust, and have close contact with consumers, which can ensure consumer loyalty and gain of new clients (Leung et al., 2013). However, companies must also be attentive to negative reviews because that can compromise the company's reputation and consumers' loyalty (Rizvi and Keole, 2015). Brunner et al. (2019) study about companies responding to negative reviews concludes that when the company gives a weak or vague response, consumers believe more in the reviews of other consumers than in the company itself. However, when the company responds with a concrete and strong response, consumers believe both in the reviews of other consumers and in the company that supplied the product. For this reason, it is important for companies to constantly monitor online comments and ensure that they can understand the consumers' opinions and perceptions.

There are currently many online platforms where customers can express their opinion about a restaurant or bar, such as Zomato and Tripadvisor. On these platforms, the biggest contributors are the consumers, who share their experiences and information about decoration, food, service, etc., rating them on a scale ranging, in general, from 1 to 5. According to Zhang et al. (2010), negative opinions have a greater impact, that is, when a customer rates a restaurant as 1 – the worst rating – compared to a customer who defines a restaurant as 5 – the best rating a customer can give. Moreover, consumers give more importance to the online platform's ratings than the official ratings, thus online ratings are becoming a more meaningful indicator for predicting a restaurant or hotel performance than traditional surveys (António et al., 2018a).

2.2 Consumer behaviour

Consumers' behaviour can be defined as the characteristics that can be identified when observing people, the consumers, making their purchases, researching, and evaluating products or services to satisfy their needs (Schiffman and Kanuk, 2009). Consumers have a decision-making process consisting of looking for information about a particular product or service, evaluating possible alternatives, and after this process, they buy the product or service. Finally, the post-purchase behaviour process (Lamb et al., 2010) reflects the intentions to buy from the same company again, as well the recommendation of the consumption of the product or service to their friends or family (Alcañiz et al., 2005). If consumers do not feel comfortable with a product or service, they quickly find a replacement to meet their need. Currently, the relationship customers have with companies must be one of responsibility, clarity, and trust (Samara and Morsch, 2005).

Information and experience sharing have, in recent times, become more techno-centric and less conventional. Consumers' behaviour is commonly influenced by prior knowledge about a specific destination, and relevant information on access and availability plays a crucial role (Ramos and Hassan, 2021).

With the growing use of the internet as a channel for the distribution of products and communication, the opportunity for interaction between organisations and consumers was created. These interactions occur when the consumer is in the search phase where the search interface is the internet (Rose et al., 2011). Understanding the mechanisms of online shopping is a priority issue for companies that compete in the online market (Constantinides, 2004).

3 Methodology

This work consists in applying a method that allows the identification, through text mining techniques, of the different factors mentioned by the users that influence the final review rating, using as case study Jamie's Oliver restaurant reviews. Generically, text mining can be viewed as a semi-automatic process, where the main objective is to extract useful patterns from unstructured textual data (Miller, 2005). The main data sources are blogs, emails, news, reports, and social networks. A document is a sequence of tokens, which can be characters or words (Calheiros et al., 2017). Text mining has become very important because we are in the era of mass communication, which makes it difficult to access unstructured information effectively and efficiently.

The methodology is divided into six steps. The first step is data preprocessing, where the text is separated into sentences and sentences into words using the following steps: word tokenisation and the removal of punctuation and stop-words is done using Natural Language Toolkit (NTLK) (Bird et al., 2009). Words are then converted to lowercase. NTLK is a set of libraries and programs used for manipulating and analysing written text, developed in Python.

The second step was to define the factors, aspects, or topics that will be analysed in the reviews, knowing that they should be related to the restaurant theme. The topics were defined based on the works of Gan and Yu (2015), which considered ‘food’, ‘service’, ‘price’, and ‘place’; of Gojali and Khodra (2016), where it was defined that the main factors were ‘food’, ‘service’, ‘price’, and ‘decor’; of Cuizon et al. (2018), which identified as topics ‘ambiance’, ‘cost’, ‘food’, ‘hygiene’, and ‘service’; and, of Luo and Xu (2019) which identified the topics ‘food’, ‘experience’, ‘value’, and ‘location’. In summary, the identified topics were the following: ‘food’, it specifies criteria for evaluating the presentation of food, ingredients, taste, and menu; ‘service’, it refers to the assistance and availability; ‘price’ or ‘cost’, it refers to customers’ evaluation on the prices of dining and service; ‘place’, it refers to where the restaurant is located; ‘decor’ or ‘ambiance’, that it is related to the decoration, atmosphere, and the general ambiance of the restaurant; ‘hygiene’, it is related to restaurant cleaning and sanitation; and ‘experience’, which refers to the service the customers received. To better understand if these topics adequately reflect the contents of the reviews, this study explored topic modelling.

The most common and popular model for topic modelling is the Latent Dirichlet Allocation (LDA) algorithm (Blei et al., 2003). The LDA model is a probabilistic, unsupervised learning algorithm that is used to generate topics, based on the terms that constitute the documents (Sendhilkumar et al., 2017). The general idea of LDA is based on the hypothesis of writing a document on one or several topics. After the creation of the model, it can be used to identify a particular topic in a text – as one term can be in more than one topic, context is important (Krestel et al., 2009). To represent the documents, lemmatised content words, namely nouns and adjectives, were used. In that sense, part-of-speech (POS) tagging was applied to the reviews and only nouns and adjectives were selected to represent the reviews. Adjectives are also important for the sentiment analysis step (see further in this section), which allows us to understand the influence of the different aspects in the final review score. POS tagging is the task of tagging each word of a sentence with the appropriate grammatical category according to the word context (Sastry et al., 2007). Lemmatisation consists of converting each word to the corresponding lexicographic entry, considering the context in which it is found.

The third step aims to identify the topics defined in the previous step in each review. The first part of the unsupervised method of Schwartz et al. (1997) was used in this study, which consists in counting how many words related to each topic appear in the text, classifying the topic by calculating the relative frequency of all words for each topic and choosing the topic with the highest frequency. This part of the work used the preprocessed data from the first step.

Subsequently, a sentiment analysis method, using a pre-trained algorithm, was used to define the sentiment of the customers with respect to each of the identified aspects (Figure 1). Sentiment analysis is the study of people's opinions about entities, individuals, topics, or events, for example. The main objective of sentiment analysis is assigning the sentiment (positive, neutral, or negative) to a specific document or phrase in an automatic way. This tool is an important method to extract opinions from unstructured documents (António et al., 2018b). The sentiment that appears in a document can be characterised in two types: explicit, which happens when the sentence directly expresses a positive or negative opinion, or implicit, where the sentence does not directly express or imply a positive or negative opinion (Liu, 2008).

Figure 1 Factors used in the study (see online version for colours)

Food : ['antipasta', 'salads', 'shrimp', 'guisada', 'panini', ...]
Price : ['costlier', 'market', 'der', 'consequence', 'gv', ...]
Location : ['terrain', 'campus', 'parkway', 'occupy', 'peasley', ...]
Service : ['tipper', 'greeter', 'catering', 'official', 'constable', ...]
Experience : ['http', 'personal', 'maternal', 'qualifies', 'uniquely', ...]
Atmosphere : ['opulent', 'scenery', 'diverse', 'erik', 'mentally', ...]

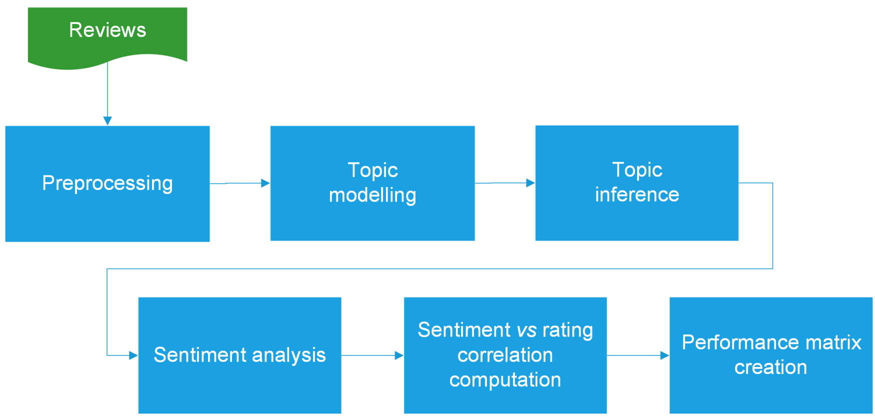
A multifunctional method was used: the method of Rolczynski (2020), which has the key concept of dividing reviews into blocks and then automatically extracting the mentioned aspects as well as their polarity. This method is based on work reported in the 2014 SemEval competition – two of these works have restaurant reviews as a case study (Rietzler et al., 2020; Sun et al., 2019; Li et al., 2019). The method consists of two steps: aspect term extraction (ATE) and Aspect Polarity Classification (APC), based on a pretrained BERT model (Devlin et al., 2019). The methods used aim to predict the exact polarity of different aspects based on context, instead of analysing the general polarity at the sentence level. This method was applied to the original sentences of the reviews, not to the preprocessed data. This method has the limitation of only processing 512 tokens each time, so it was necessary to separate the sentences into 512 tokens. In case sentences exceeded this length, to guarantee that a sentence was not separated in the middle, a sentence-breaking condition was set – before the 512 tokens, break on the closest punctuation (only the period, exclamation, and question mark were considered, since these are the punctuations that indicate the end of a sentence); if no punctuation was found on the 512 tokens, the sentence is broken on the closest space. This method returns three probabilities: the probability that the sentiment is negative, neutral, or positive. The overall sentiment represented in the sentence was the one that had the highest value, converting the positive sentiment to the value of 1, the neutral sentiment to the value of 0, and the negative sentiment to the value of –1.

To calculate the correlation between the review rating and the sentiment associated with each category, in order to understand the relevance and impact of each category on the final rating, the linear correlation statistical method was applied. To calculate the correlation the standard Pearson's coefficient was used, or linear correlation, which shows the correlation between variables, with a coefficient between –1 and 1.

The last step consisted in creating a performance matrix, to compare the correlation values between the categories and the final rating, and sentiment of each category.

The performance matrix was based on the work of Ferreira and Fernandes (2015). The matrix has four quadrants and compares customer satisfaction with a specific topic and the importance that the customer places on that same topic. This analysis allows, through a representation based on Cartesian coordinates, to identify areas where a company should focus, reduce, or maintain its efforts and also where there are the greatest deviations between what is important to the customer and what they are receiving/feeling. Figure 2 presents the complete sequence of steps.

Figure 2 Methodology (see online version for colours)



4 Analysis and discussion of results

4.1 Sample characterisation

There were 499 reviews selected from Zomato (5%) and TripAdvisor (95%). 80% of the reviews were from the England restaurant and 20% from the Portugal restaurant. The time period considered is from January 2015 until December 2020. In this extraction, only reviews written in English were used. The data consist of 499 reviews, 3173 sentences, and 47413 words in total. The largest review consists of 123 words, and the smallest has only 1 word. Each review, on average, consists of 6.35 sentences and 94.83 words.

4.2 Model implementation

The LDA topic model was applied in order to validate the relevant topics for this study, according to the topics identified in the works of Gan and Yu (2015), Gojali and Khodra (2016), Cuizon et al. (2018), and Luo and Xu (2019), being these topics: 'food'; 'service'; 'price' or 'cost'; 'place'; 'decor' or 'ambiance'; 'hygiene' and 'experience'.

The application started with seven topics, where it was possible to identify all topics except 'hygiene'. To evaluate the created model, a topic coherence model (Roder et al., 2015) was used. On the London restaurant dataset the model achieved a coherence of 0.42. Using the data from the Lisbon restaurant, a coherence of 0.54 was obtained.

As only six topics were identified, the LDA model was applied again with only six topics and the coherence of the London restaurant dataset increased to a value of 0.55, where the coherence with the Lisbon restaurant dropped to 0.41. Through the results obtained, it was possible to identify the following six topics: ‘food’; ‘service’; ‘price’ or ‘cost’; ‘place’; ‘decor’ or ‘ambience’; and ‘experience’, labelled in Figures 3 and 4, as ‘food’, ‘service’, ‘price’, ‘location’, ‘atmosphere’ (representing ‘ambience’ and ‘ decor’) and ‘experience’.

Figure 3 LDA Lisbon (see online version for colours)

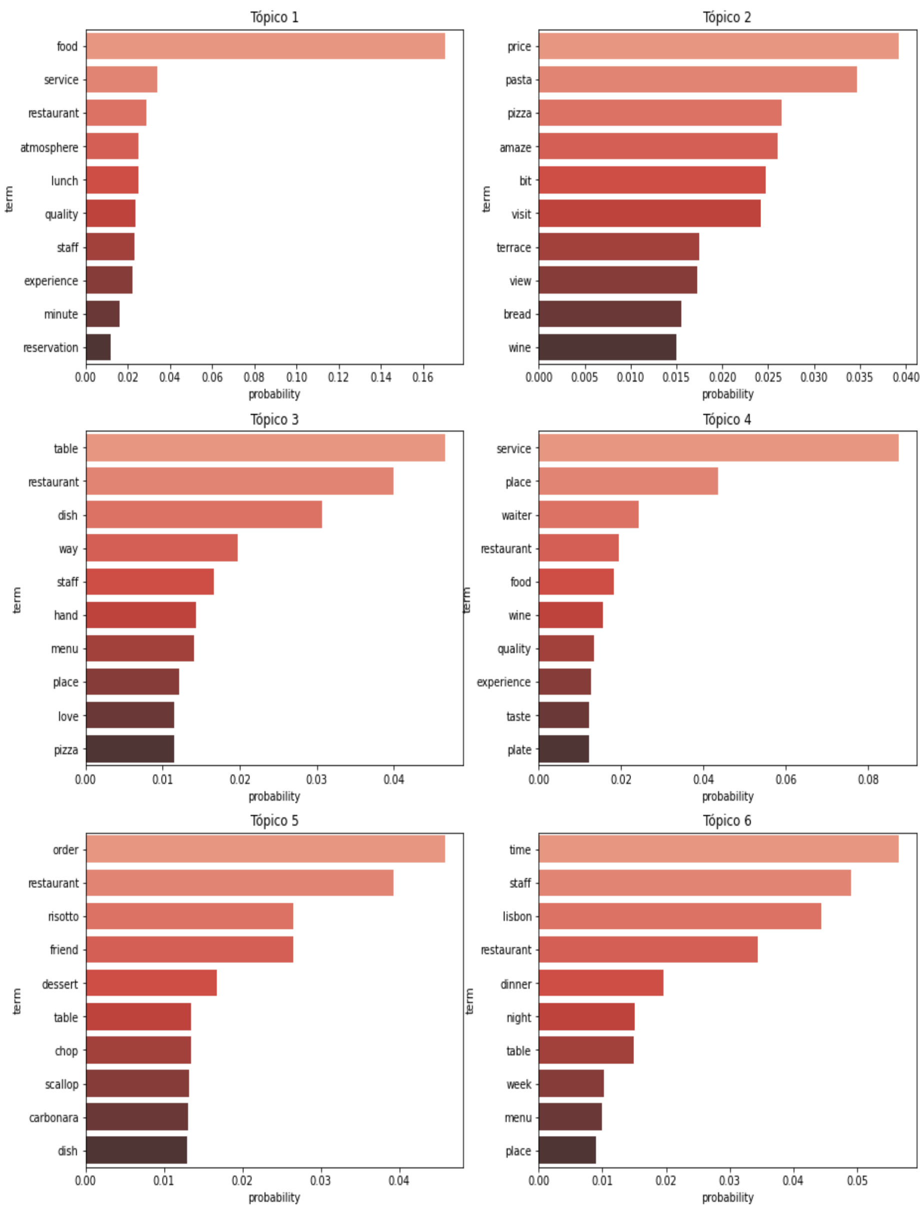
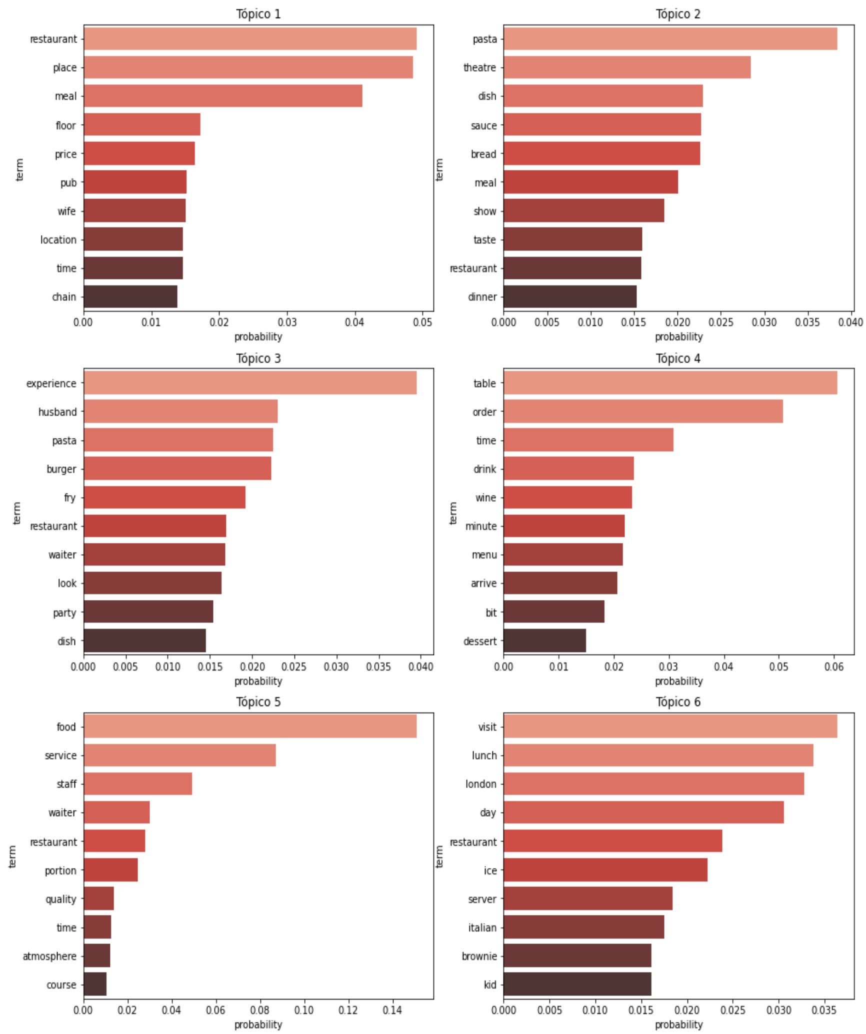


Figure 4 LDA London (see online version for colours)



In the third step, the goal was to identify the topics defined in the previous step in each of the reviews. To achieve this, as previously mentioned, the unsupervised method of Schwartz et al. (1997) was considered.

Of the 542 sentences from restaurant reviews in Portugal, the model could not automatically identify the topic in 33. 256 sentences were manually analysed to evaluate the method's performance. To evaluate the performance, it was used the accuracy. This method achieved a value of 77%. After this validation, the following improvements were applied:

- The word 'dessert' was added to the food library since it was a common spelling mistake to write 'dessert' with just one 's';
- If the number of words was the same for two categories and this number was the maximum (a tie), the sentence was duplicated and both categories were considered.

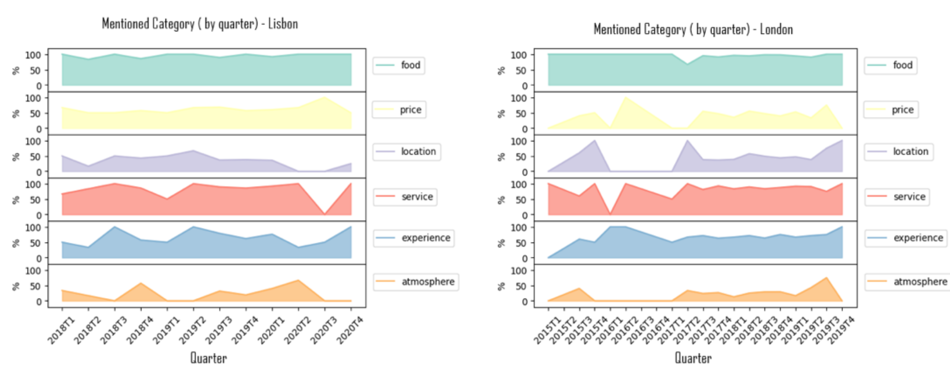
After these improvements, 256 sentences were randomly selected to ensure that the improvements were not biased by the previous validation set, and an accuracy of 86% was obtained. The method identified that 20% of the sentences had more than one category mentioned. This method could not automatically identify the category in 4% of the sentences in the London and Lisbon reviews, and these were then identified by hand to have 100% of the data collected for the next steps.

4.3 Results analysis

The most mentioned factor in the reviews of the Lisbon restaurant is ‘food’, being mentioned in 94% of the reviews, followed by ‘service’, with 86%, and ‘experience’, with 66%, and the least mentioned factors were ‘atmosphere’ and ‘location’. Similar to the factors observed in the Lisbon results, the most mentioned factors in the London restaurant were ‘food’, ‘service’, and ‘experience’, while the least mentioned were ‘atmosphere’ and ‘location.’

After analysing the percentage of categories mentioned in general, this same metric was analysed divided by quarters (Figure 5). It was performed a relative analysis since the number of reviews is different throughout the months. It was found that ‘food’ is a factor that remains constant over time, both in the Lisbon and London restaurants, and the ‘atmosphere’ factor is mentioned more after 2019. It was also possible to verify that in the Lisbon restaurant after the second quarter of 2020, the mentions of the factors ‘service’, ‘location’, and ‘experience’ decreased.

Figure 5 Mentioned category (by quarter) (see online version for colours)



The results from sentiment analysis, the fourth step, showed that the factor with the most positive sentiment, both in the Lisbon restaurant and in the London restaurant, is ‘atmosphere’, with the values of 0.81 and 0.23, respectively, where the predominant sentiment in the Lisbon restaurant is 1. The next most positive sentiment in the London restaurant is in the ‘food’ factor, with 0.12, and 0.35 in the Lisbon restaurant. The second most positive sentiment for the Lisbon restaurant is the service, with 0.45. For the London restaurant, the estimated value is 0.05, a neutral sentiment. For the ‘location’ factor, the sentiment in Lisbon was 0.15 and in the London restaurant was 0.09. The ‘experience’ factor has a practically neutral sentiment for both restaurants, but it can be seen that the restaurant in Portugal initially showed a more negative sentiment that increased over time, while the London restaurant shows the opposite phenomenon.

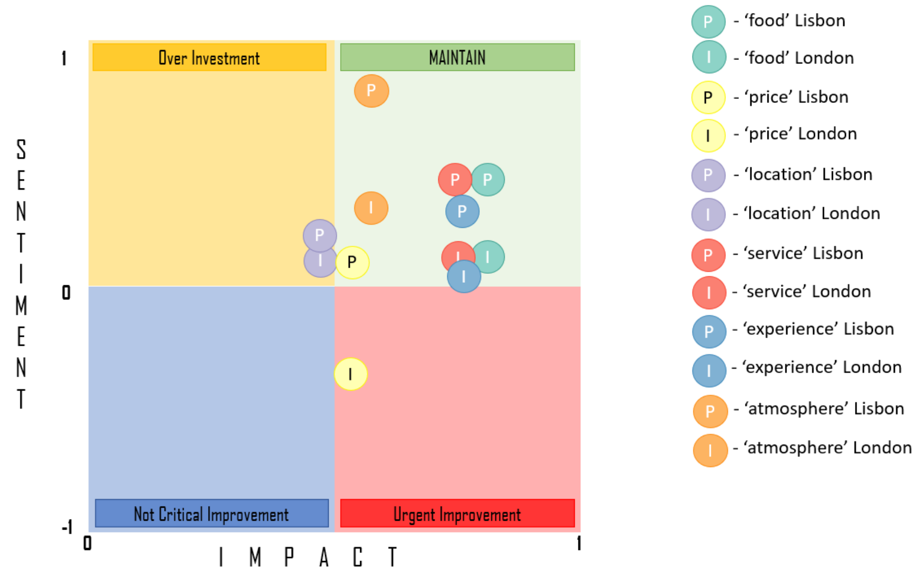
Finally, the factor 'price' is the one that showed the biggest difference between both restaurants: the sentiment in Lisbon is approximately neutral, while the sentiment in London is negative. However, in the Lisbon restaurant, it can be seen that after the second quarter of 2020, this aspect was more mentioned, but the sentiment started to be more negative (Figure 6). In general, it can be seen that the Lisbon restaurant sentiment score is more positive than the one of the London restaurant.

Figure 6 Category sentiment analysis (by quarter) (see online version for colours)

The fifth step was to verify the correlation between the sentiment of each factor and the final rating. To observe this correlation, the Lisbon and London reviews were analysed together. The correlation between sentiment and rating, between rating and number of tokens, and also the correlation of sentiment between factors.

A performance matrix was created to compare the values of the correlation between the factors, the final rating, and the sentiment for each factor. In Figure 7, the x-axis represents the impact of the factor, that is, the correlation between the rating and the factor, and the y-axis represents the average sentiment for each factor. This analysis assumes that there is a linear relationship between importance and performance. The 'Urgent Improvement' quadrant shows the topics that have the most impact on the final rating, but where the sentiment is rated as negative. To increase customer satisfaction, it is crucial that the company focuses on that factor – this is the case for the 'price' factor in the London restaurant. The 'Maintain' quadrant refers to topics that have a large impact on the final rating and where the sentiment is evaluated as positive, thus representing the attributes that the restaurant should maintain competitive advantage. It can be observed that this is where most of the factors are, both in the Lisbon and London restaurants, but it is possible to verify that there are three factors in the London restaurant, 'service', 'food', and 'experience', which are already close to the 'Urgent Improvement' quadrant. The 'Not-critical improvement' quadrant represents the topics that are considered less relevant and where the sentiment is negative. It is usually not necessary to focus on these attributes. There are no factors in this quadrant. The last quadrant, 'Over-investment', presents the topics where the sentiment is positive but where the impact on the rating is not mentioned as important, meaning that the resources that are being invested in these factors could be better used on other topics, this is the case of 'location' factor in both restaurants.

Figure 7 Performance matrix (see online version for colours)



5 Conclusion

This work focused on the application of a method that allows the identification, through text mining, of the different aspects mentioned in users' reviews that influence the final review rating, using as case study two specific Jamie's Oliver restaurants reviews.

The first conclusion is that the aspects (or factors) most mentioned in the reviews were 'food' and 'service', and the least mentioned were 'atmosphere' and 'location'. Analysing the results, it was found that in the Lisbon restaurant, from the second quarter of 2020, the mentions of the factors 'service', 'location', and 'experience' decreased. This can be justified by the fact that restaurants were closed due to the pandemic and customers only consumed the restaurant's meals by take-out. Linked to this point is the fact that the factor 'price' was mentioned more often in the second quarter of 2019 and the sentiment had a negative trend – since customers were not enjoying the atmosphere and service, they were not willing to pay the same price. In summary, from the analysis it is possible to conclude that the sentiment in each of the factors is more positive in the Lisbon restaurant than in the London restaurant, so the overall customer satisfaction is higher in the Lisbon restaurant. This conclusion can also be drawn from the restaurant's average rating, since the Lisbon restaurant has an average of 3.8 points, while the London restaurant has 3 points. The 'price', 'food', 'service', and 'experience' factors can be considered the weaknesses of London restaurant and possibly contributed its closure. Regarding the strength of the correlation between the factors and the final rating, it was concluded that all factors have a strong correlation with the final rating, highlighting the factor 'food' with a correlation of 0.73. The correlation between factors was also analysed, concluding that there is a strong correlation between all of them, excluding the correlation between 'price' and 'location'. The strongest correlation is the correlation between 'experience' and 'price', with a value of 0.66.

At the end of the results analysis, it was possible to verify that the highest correlation between the final rating and the sentiment concerns the factor ‘food’. It was also verified that from the year 2018 there is an increase in the mentions of the factor ‘ambiance’. Finally, the percentage of mentions of ‘price’ is seen to be higher in the Lisbon restaurant than in the London restaurant. However, considering the ranking of restaurants related to the price ranking, taking into account other restaurants in the same location, it was verified that the restaurant in Portugal is in the top of the most expensive restaurants, while the London restaurant is not. However, if we check the sentiment related to the factor ‘price’, it is more negative in the London restaurant than in the Lisbon restaurant.

One of this study’s limitations was the sample of reviews that was small. Another limitation was segmenting the reviews into sentences, which sometimes led to the context or the true sentiment being lost. In 4% of the total sentences, the method could not automatically identify the corresponding factor. Finally, there was also a linguistic limitation, since there are few text mining tools in other languages than English.

For future research, it is proposed to apply this model through the creation of an app or an add-on for Zomato or Tripadvisor platforms, where the results would be generated automatically and in real time, adding points such as the positioning of the restaurant in relation to nearby restaurants or restaurants with the same type of cuisine.

In terms of the methods used, other approaches may be used: instead of classic topic modelling step, a clustering methodology based on contextual embeddings could be used; in the topic classification, approaches based on semantic similarity could improve the match between the topics and the reviews, and for the last model step, when the impact of each factor is analysed, other statistical methods may be used, such as the correlation coefficient, which calculates the relationship of a quantitative variable with another nominal variable.

The model introduced allows to identify the individual sentiment by factor and to identify how each factor impacts the final rating; it helps to understand the restaurant’s performance in each factor over time, validating the points to be improved, those that have to maintain or reduce the investment, or which an improvement action is urgent. This model added in platforms such as Zomato or TripAdvisor could contribute to the growth of restaurants or hotels and help understanding which factors contribute the most to customer satisfaction.

Acknowledgement

This work was undertaken at ISTAR-Information Sciences and Technologies and Architecture Research Center from Iscte-Instituto Universitário de Lisboa (University Institute of Lisbon), and INESC-ID Lisboa, and it was partially funded by the Portuguese Foundation for Science and Technology (Projects ‘FCT UIDB/04466/2020’ and “FCT UIDB/50021/2020”).

References

Alcañiz, E.B., García, I.S. and Blas, S.S. (2005) ‘Relationships amog residents’ image, evaluation of the stay and post-purchase behaviour’, *Journal of Vacation Marketing*, Vol. 11, No. 4, pp.291–302, doi: 10.1177/1356766705056626.

António, N., Almeida, A.M., Nunes, L., Batista, F. and Ribeiro, R. (2018a) 'Hotel online reviews: creating a multi-source aggregated index', *International Journal of Contemporary Hospitality Management*, Vol. 30, No. 12, pp.3574–3591, doi: 10.1108/IJCHM-05-2017-0302.

António, N., Almeida, A.M., Nunes, L., Batista, F. and Ribeiro, R. (2018b) 'Hotel online reviews: different languages, different opinions', *Information Technology and Tourism*, Vol. 18, No. 1, pp.157–185, doi: 10.1007/s40558-018-0107-x.

Bird, S., Loper, E. and Klein, E. (2009) *Natural Language Processing with Python*, O'Reilly Media Inc., Sebastopol, CA, USA.

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

Brunner, C.B., Ullrich, S. and De Oliveira, M. (2019) 'The most optimal way to deal with negative consumer review: can positive brand and customer responses rebuild product purchase intentions?', *Internet Research*, Vol. 29, No. 1, pp.104–122, doi: https://doi.org/10.1108/IntR-08-2017-0307

Calheiros, A.C., Moro, S. and Rita, P. (2017) 'Sentiment classification of consumer-generated online reviews using topic modelling', *Journal of Hospitality Marketing and Management*, Vol. 26, No. 7, pp.675–693, doi: 10.1080/19368623.2017.1310075.

Constantinides, E. (2004) 'Influencing the online consumer's behavior: the web experience', *Internet Research*, Vol. 4, No. 2, pp.111–126, doi: 10.1108/10662240410530835.

Constantinides, E. and Fountain, S.J. (2008) 'Web 2.0: conceptual foundations and marketing issues', *Journal of Direct, Data and Digital Marketing Practice*, Vol. 9, No. 3, pp.231–244.

Cuizon, J.C., Lopez, J. and Jones, D.R. (2018) 'Text mining customer reviews for aspect based restaurant rating', *International Journal of Computer Science and Information Technology*, Vol. 10, No. 6, pp.43–52.

Devlin, J., Chang, M., Lee, K. and Toutanova, K. (2019) 'BERT: pre-training of deep bidirectional transformers for language understanding', *Proceedings of 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, Minneapolis, USA, pp.4171–4186.

Dubey, A.K. (2016) 'Future technology and service industry: a case study of travel and tourism industry', *Global Journal of Enterprise Information System*, Vol. 8, No. 3, pp.65–69.

Ferreira, A.D. and Alturas, B. (2010) 'Social media no contexto das empresas/Marcas: uma proposta de investigação [Social media in the context of companies/Brands: a research proposal]', *Proceedings of the IASK International Conference InterTIC 2010*, Oviedo, Spain, pp.292–294.

Ferreira, H. and Fernandes, P.O. (2015) 'Importance-performance analysis applied to a laboratory supplies and equipment company', *Conference on ENTERprise Information Systems/International Conference on Project Management/Conference on Health and Social Care Information Systems and Technologies,* Bragança, Portugal, pp.824–831.

Gan, Q. and Yu, Y. (2015) 'Restaurant rating: industrial standard and word-of-mouth A text mining and multi-dimensional sentiment analysis', *Proceedings of the 48th Hawaii Internacional Conference on System Sciences,* USA, pp.1332–1340, doi: 10.1109/HICSS.2015.163.

Gojali, S. and Khodra, M.L. (2016) 'Aspect based sentiment analysis for review rating prediction', *2016 International Conference On Advanced Informatics: Concepts, Theory and Application* (*ICAICTA*), Penang, Malaysia, pp.1–6, doi: 10.1109/ICAICTA.2016.7803110.

Krestel, R., Frankhauser, P. and Nejdl, W. (2009) 'Latent dirichlet allocation for tag recommendation', *Proceedings of the Third ACM Conference on Recommender Systems (RecSys'09)*, New York, USA, pp.23–25.

Lamb, C.W., Hair, J.F. and McDaniel, C.D. (2010) *Marketing 3*.0, Mason.

Leung, D., Law, R., van Hoof, H. and Buhalis, D. (2013) 'Social media in tourism and hospitality: a literature review', *Journal of Travel and Tourism Marketing*, Vol. 30, Nos. 1–2, pp.3–22.

Li, X., Bing, L., Zhang, W. and Lam, W. (2019) 'Exploiting BERT for end-to-end aspect-based sentiment analysis', *Proceedings of the 2019 EMNLP Workshop W-NUT: The 5th Workshop on Noisy User-generated Text*, Hong Kong, pp.34–41.

Liu, B. (2008) *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, 2nd ed., Springer Berlin Heidelberg.

Luo, Y. and Xu, X. (2019) 'Predicting the helpfulness of online restaurant reviews using different machine learning algoritms: a case study of yelp', *Sustainability*, Vol. 11, No. 19, p.5254, doi: https://doi.org/10.3390/su11195254

Miller, T. (2005) *Data and Text Mining: A Business Applications Approach*, Pearson Education International, Upper Saddle River, USA.

Park, Y. and Gretzel, U. (2007) 'Success factors for destination marketing web sites: a qualitative meta-analysis', *Journal of Travel Research*, Vol. 46, No. 1, pp.46–63.

Parreira, C., Fernandes, A.L. and Alturas, B. (2021) 'Digital tourism marketing: case study of the campaign can't skip Portugal', *Smart Innovation, Systems and Technologies*, Vol. 205, No. 1, pp.759–768, doi: 10.1007/978-981-33-4183-8_61.

Ramos, C.M.Q. and Hassan, A. (2021) 'Evaluation of the tourist experience using a framework for the analysis of the influence of social media on tourist behaviour', *International Journal of Tourism Policy*, Vol. 11, No. 3, pp.207–221.

Rietzler, A., Stabinger, S., Opitz, P. and Engl, S. (2020) 'Adapt or get left behind: domain adaptation through BERT language model finetuning for aspect-target sentiment classification', *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille, France, pp.4933–4941.

Rizvi, N.T.S.H. and Keole, R. (2015) 'A preliminary review of web-page recommendation in information retrieval using domain knowledge and web usage mining', *International Journal of Advance Research in Computer Science and Management Studies*, Vol. 3, No. 1, pp.156–166.

Roder, M., Both, A. and Hinneburg, A. (2015) 'Exploring the space of topic coherence measures', *Proceedings of Eighth ACM International Conference on Web Search and Data Mining,* pp.399–408, doi: https://doi.org/10.1145/2684822.2685324

Rolczynski, R. (2020) *Aspect-Based-Sentiment-Analysis 2.0.2*, Online at Pypi.org: https://pypi.org/project/aspect-based-sentiment-analysis/

Rose, S., Hair, N. and Clark, M. (2011) 'Online customer experience: a review of the business-to-consumer online purchase context', *International Journal of Management Reviews*, Vol. 13, No. 1, pp.24–39.

Samara, B.S. and Morsch, M. (2005) *Comportamento do consumidor: conceitos e casos [Consumer Behavior: Concepts and Cases]*, Prentice Hall, São Paulo, Brazil.

Sastry, G., Chaudhuri, S. and Reddy, P. (2007) 'An HMM based part-of-speech tagger and statistical chunker for 3 Indian languages', *IJCAI*-07, Hyderabad, India. pp.13–16.

Schiffman, L.G. and Kanuk, L.L. (2009) *Consumer Behavior*, PrenticeHall, Englewood Cliffs, USA.

Schwartz, R., Imai, T., Kubala, F., Nguyen, and Makhoul, J. (1997) 'A maximum likehood model for topic classification of broadcast news', *5th European Conference on Speech Communication and Technology*, Rhodes, Greece, pp.1455–1458, doi: 10.21437/Eurospeech.1997-422.

Sendhilkumar, S. and Srivani, M. and, Mahalakshmi, (2017) 'Generation of word clouds using document topic models', *Second International Conference on Recent Trends and Challenges in Computational Model*, Tindivanam, Tamilnadu, India, pp.306–308, doi: 10.1109/ICRTCCM.2017.60.

Sun, C., Huang, L. and Qiu, X. (2019) 'Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence', *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* Minneapolis, Minnesota, USA, pp.380–385.

Tripadvisor (2020) *Acerca Do Tripadvisor*, 30 June, Obtido de Tripadvisor: https://tripadvisor.mediaroom.com/pt-about-us

Tripadvisor (2020) *Jamie's Italian, Lisbon, Portugal*, 30 July, Obtido de Tripadvisor: https://www.tripadvisor.pt/Restaurant_Review-g189158-d13473338-Reviews-Jamie_s_Italian_Lisboa-Lisbon_Lisbon_District_Central_Portugal.html

Tripadvisor (2020) *Jamie's Italian, London, England* 27 July. Obtido de Tripadvisor: https://www.tripadvisor.pt/Restaurant_Review-g186338-d6214093-Reviews-Jamie_s_Italian-London_England.html

Zhang, J.Z. and Chang, C. (2021) 'Consumer dynamics: theories, methods, and emerging directions', *Journal of the Academy of Marketing Science*, Vol. 49, No. 1, pp.166–196, https://doi.org/10.1007/s11747-020-00720-8

Zhang, Z., Ye, Q., Law, R. and Li, Y. (2010) 'The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews', *International Journal of Hospitality Management*, Vol. 29, No. 4, pp.694–700.

Zomato (2020) *Jamie's Italian, Principe Real, Lisboa*, 30 July, Obtido de Zomato: https://www.zomato.com/pt/grande-lisboa/jamies-italian-pr%C3%ADncipe-real-lisboa

Zomato (2020) *Jamie's Italian, Soho, London*, 27 July, https://www.zomato.com/pt/london/jamies-italian-soho