Expert Systems With Applications 184 (2021) 115546

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Supporting digital content marketing and messaging through topic modelling and decision trees

Andreas Gregoriades [a,*,1], Maria Pampaka [b,2], Herodotos Herodotou [a,3], Evripides Christodoulou [a]

[a] Cyprus University of Technology, Limassol, Cyprus
[b] The University of Manchester, Manchester, UK

ARTICLE INFO

Keywords:
Topic modelling
Cultural and economic distance
Decision trees
Shapley additive explanation
Tourists' reviews

ABSTRACT

This paper presents a machine learning approach involving tourists' electronic word of mouth (eWOM) to support destination marketing campaigns. This approach enhances optimisation of a critical aspect of marketing campaigns, that is, the communication of the right content to the right consumers. The proposed method further considers aggregate cultural and economic-related information of the tourists' country of origin with topic modelling and Decision Tree (DT) models. Each DT addresses different dimensions of culture and purchasing power and the way these dimensions are associated with the topics discussed in eWOM, thus revealing patterns relating tourists' experiences with potential explanations for their dissatisfaction/satisfaction. The method is implemented in a case study in the context of tourism in Cyprus focusing on two hotel groups (2/3 and 4/5 stars) to account for their differences. Patterns emerged from the extraction of rules from DTs illuminate combinations of variables associated with tourist experience (negative or positive) for each of the two hotel categories and verify the asymmetric relationship between service performance and satisfaction. The approach can be used by management during marketing campaigns to design messages to better address the desires and needs of tourists from different cultural and economic backgrounds, as these emerge from the data analysis.

1. Introduction

Digital content marketing (DCM) and messaging are effective, yet risky methods to target customers during a marketing campaign since a wrong message may have negative effects. DCM as an e-marketing approach, utilises contemporary artificial intelligence (AI) techniques and attempts to optimize consumers' experience with a product or service, while maintaining a profit (Pulizzi, 2014), via in-depth understanding of the target consumers' needs. However, contacting customers with a marketing message entails a cost, so marketers use segmentation and targeting to focus on customers who are more likely to respond and be satisfied after experiencing a service.

In tourism, DCM plays an important role in shaping consumers' decisions. Retrospective information regarding tourists' satisfaction or dissatisfaction with a service derived from electronic word of mouth (eWOM) is important when designing DCM communications, to effectively utilise a service's strengths and address its weaknesses. Touristic destinations aim to provide a positive experience for visitors with accommodation playing a key role to this. Popular methods to assess satisfaction with accommodation include consumer feedback via eWOM or surveys; the former represents unstructured reports of customers' personal experience that influence consumers decision making (Dwivedi et al., 2020), and is popular with machine learning analytical techniques. EWOM has been employed extensively in various domains due to its ability to shape consumers' opinions or to support the evaluation/design of marketing campaigns and strategies. For example, research in tourism has considered internal (Guo, Barnes, & Jia, 2017) or external information in regards to eWOM content to explore the factors that lead to eWOM behaviour (e.g., Yen & Tang, 2019); internal factors refer to the actual eWOM content whilst external factors refer to other features

* Corresponding author.
E-mail addresses: andreas.gregoriades@cut.ac.cy (A. Gregoriades), maria.pampaka@manchester.ac.uk (M. Pampaka), herodotos.herodotou@cut.ac.cy (H. Herodotou), ep.xristodoulou@edu.cut.ac.cy (E. Christodoulou).
[1] ORCID: 0000-0002-7422-1514.
[2] ORCID: 0000-0001-5481-1560.
[3] ORCID: 0000-0002-8717-1691.

https://doi.org/10.1016/j.eswa.2021.115546
Received 14 August 2020; Received in revised form 16 June 2021; Accepted 1 July 2021
Available online 6 July 2021
0957-4174/

related to the eWOM such as demographic information of eWOM's producer.

During a marketing campaign, consumers' needs, wants, and values must be considered effectively, so as the message can motivate the consumer, given evidence that motivation is a major driver in consumer decision making and expectation formation (e.g., Park & Nicolau, 2015), with eWOM constituting a valuable source of information for such analysis. According to Cognitive fit theory (Vessey, 1991), designing effective content for marketing communication requires matching consumers' information needs, which are linked to consumers' wants, with what and how information is presented in a message. When this fit is optimum, it impacts consumers' decision making towards purchasing a product or service. An important parameter in improving this fit is to consider recipients' culture. Evidence suggests that culture and purchasing power are linked to tourist satisfaction (Morgeson, Mithas, Keiningham, & Aksoy, 2011); culture was associated with individuals' wants (Kotler & Armstrong, 2018), and purchasing power with service satisfaction (Blut, Teller, & Floh, 2018). When visitors interact with different cultural and economic contexts than their own (cultural and economic distance), their culture shapes their experience (Sharpley, 2014) by affecting their perception of service, which in turn is expressed in eWOM. As measuring culture at the individual personal level is challenging, national scores denoting general/broader cultural aspects, such as those of Hofstede (Swierstra & Rip, 2007), are used extensively in the literature.

Most previous work in machine-learning-based DCM utilise structured data from search engines, and consumers' engagement with websites and social networks (Aviera, Akinwale, & Fontecha, 2021), to extract consumers' demographics, behaviours or user details (names etc.), for better targeting. Such methods require the integration of data from several sources, which may not be always accessible or their potential access could involve privacy concerns that could lead to rejection of the message (Winter, Maslowsk, & Vosc, 2020). Moreover, given the influence of eWOM on consumers decision making, past research examined mostly how DCM influences eWOM and little on how insights from eWOM analytics can assist the design of messages for DCM (Dwivedi et al., 2020). Additionally, AI-based consumer targeting methods focus mostly on "which" content and "when" to show it to customers using structured contextual or user-based data, with limited work on "what" information to include in the content. This paper aims to address this research gap, by focusing on the integration of unstructured textual data (i.e., consumers' opinions from eWOM) with information on the culture and economic conditions of eWOM's authors, given their link to satisfaction, to assist in the design of messages that will motivate target consumers during a campaign. Building on the cultural and economic distance theory(Ahn & McKercher, 2015), our proposed method uses topic modelling and decision trees (DTs) to identify patterns that could support DCM with appropriate content. The method automatically identifies eWOM's topics that relate to satisfaction, and subsequently associates these with tourists' cultural and economic distance to the host country. The effect of eWOM's topics on satisfaction, cultural and economic dimensions is modelled as binary classification utilising DT to increase results' interpretability. The application of DTs is motivated by their explainability property (Voosen, 2017) and their intuitive capacity to visualise the relationships between variables.

This paper builds on our previous work on eWOM analysis for product positioning evaluation using Latent Dirichlet Allocation (LDA) (Gregoriades & Pampaka, 2020) and applies the Structural Topic Modelling (STM) (Gambhir & Gupta, 2017) approach that extends established topic modelling with metadata. The paper also extends the analysis of the joint effect of culture and purchasing power on hotel guests satisfaction/dissatisfaction using DTs (Christodoulou, Gregoriades, Pampaka, & Herodotou, 2020), with a novel investigation of the effect of cultural distance and purchasing power difference of tourists against the host country on eWOM, for two hotel groups (2/3 and 4/5 stars), which can subsequently inform DCM about content that matches cultural and economic features of target tourist groups. The paper makes the following key contributions:

1. We propose a practical and explainable methodology for extracting patterns relating to tourists' experiences and potential reasons for dissatisfaction/satisfaction from eWOM, using the cultural distance theory and economic difference, to support message design in DCM.
2. The proposed methodology combines different techniques to optimise the DT performance, using an investigation of various feature selection methods, imbalance data treatment analysis, exploiting multiple DTs per problem dimension and aggregate insight analyses.
3. We implement our methodology in a real case study in the context of tourism in Cyprus focusing on two hotel groups (2/3 and 4/5 stars) and verify the asymmetric relationship between service performance and satisfaction.
4. The resulting patterns are validated against domain knowledge and the results of the DT approach are compared against most prevalent white box techniques.

The paper is organized as follows. The next section reviews the literature on assessment of customer satisfaction using eWOM and the effect of cultural and purchasing power factors on the experience of tourists at destinations' hotels. The proposed methodology for supporting DCM for campaign management is then overviewed, followed by a case study implementing the method on data from Cyprus and describing the results. At the concluding section, the results are compared against domain knowledge and other white box techniques and discussed along with their implications for management in the sector. Limitations and future research directions are also addressed.

2. Literature review

2.1. Assessing customer satisfaction to inform DCM

Central to DCM is getting the right message to the right consumer by understanding customer needs. DCM in contrast to advertisement that try to persuade consumers to buy, focuses on providing consumers with content that is valuable, and thus aims to develop consumer engagement, trust, and relationships, to cultivate sales indirectly (Hollebeek & Macky, 2019). A valuable resource for attaining this goal is information from eWOM and other customer related data (Dwivedi et al., 2020), which can facilitate targeting and personalization in marketing communication (Winter et al., 2020). Popular eWOM analysis techniques relating to satisfaction include sentiment analysis (SA) and topic modelling for the identification of themes discussed in eWOM. Example applications of these are in service-design (Dwivedi et al., 2020) and touristic destination marketing (Kirilenko, Stepchenkova, & Dai, 2021). The ability of SA to measure emotions through algorithms that detect polarity (Pang & Lee, 2008) makes it an attractive method for eWOM analysis and a valuable tool in DCM. From the three main approaches for SA (i.e., Machine Learning, Lexicon-based Methods, and Linguistic Analysis techniques), Machine Learning is considered the simplest and most effective and its most popular implementations are with Naïve Bayes and Support Vector Machines (Boiy & Moens, 2009).

In hotel evaluations, star rating is commonly used as an indication of customers' satisfaction. However, such ratings could be incorrect (Zhang, Ye, Zhang, & Li, 2011), with a rating being high but the actual review negative and vice versa. Hence, when analysing reviews, Valdivia, Luzon, and Herrera (2017) suggest avoiding the user rate as a label of sentiment for the whole review but instead analyse the opinions mentioned in the review in depth. Furthermore, star ratings are aggregate measures that do not always capture the complexity of a review, whereby certain sub-parts might criticise a service while other parts may glorify an asset of a hotel. In such cases, reviews are most of the time neutral on an aggregate level since most people are reluctant to make negative reviews. Past evidence (Koh, Hu, & Clemons, 2010) also suggests that reviewers avoid giving low hotel scores unless they had a very negative experience. This suggests that review scores alone could be a biased indicator of satisfaction. In response to these limitations, the proposed method uses sentiment classification in addition to reviewer's ratings to assess the tone of the review and measure its implied sentiment to inform tourist satisfaction.

There are several theoretical frameworks to measure satisfaction. A prominent one is the 3-factor theory that classifies service attributes into three categories: basic factors (minimum requirements that cause dissatisfaction if not fulfilled but do not lead to satisfaction if fulfilled or exceeded; e.g., non-smelly room), performance factors (lead to satisfaction if performance is high and to dissatisfaction if performance is low; e.g., polite staff), and exciting factors (increase customer satisfaction if delivered but do not cause dissatisfaction if they are not delivered; e.g., made feel special) (Matzler, Bailom, Hinterhuber, Renzl, & Pichler, 2004). DTs are used in this work since they enable expressing non-linear relationships between variables such as the asymmetric relationship between satisfaction and performance of a service (Busacca & Padula, 2005), which usually characterizes accommodation services: some service attributes impact satisfaction positively or negatively in a non-uniform manner when their performance is high or low, respectively.

Most satisfaction evaluation methods, however, follow a top-down approach using questionnaires with pre-specified factors of satisfaction to be evaluated. However, surveys have been criticised as being time consuming and expensive. A bottom-up approach enables identifying factors that are associated with satisfaction from big data and has several advantages concerning data quality in comparison to surveys. Reviews are considered more objective because they are spontaneous (Schuckert, Liu, & Law, 2015). Consequently, analysis of eWOM from reviews websites has slowly become a mainstream approach for evaluating satisfaction in hospitality and tourism (Guo et al., 2017). However, despite the benefits from utilising eWOM, there is limited research on methods that harness the unstructured part (eWOM) of reviews for consumers segmentation and targeting (Liu, Liao, Huang, & Liao, 2019) with most work utilising the structured part of eWOM (rating score). This work contributes in this direction by identifying associations between topics discussed in eWOM with satisfaction. Such insights can provide valuable input to DCM.

2.2. Culture and satisfaction

Culture is found to be linked to peoples' wants and needs at a coarse grain level (Morgeson et al., 2011) with Songshan and John (2019) reporting that certain dimensions of culture are positively related to visitor satisfaction and others negatively. Most studies focusing on culture utilise survey methods for data collection, with only limited work using eWOM.

There are several well-developed national culture models with Hofstede's model (Hofstede, 1980; Hofstede, Hofstede, & Minkov, 2010) being one of the most popular ones, and hence adopted in this study. The model distinguishes between six different traits of culture:

(1) *Power Distance:* the trait based on which people accept a higher degree of unequally distributed power;
(2) *Individualism:* the trait of prioritizing ones' own self and their immediate families. Its opposite is collectivism, where people feel stronger belonging in a group, as in the example of Japan;
(3) *Masculinity:* refers to a preference for material rewards, assertiveness, achievement and heroism. Its opposite is femininity, where cooperation, modesty, and caring for the weak is preferred. We acknowledge that these connotations with regards to masculinity and femininity are stereotypical and are vastly criticised by scholars in other fields, however their use here and within this framework is solely intended to capture this 'gender inequality' aspect of a culture.
(4) *Uncertainty Avoidance:* refers to people feeling threatened by uncertain or unknown situations;
(5) *Long-term Orientation:* refers to the trait of perseverance, thrift, and adapting to changing circumstances. Its opposite refers to a preference for stability and respect for tradition. Examples of countries that score high in long-term orientation is China and other East Asian countries;
(6) *Indulgence:* refers to the tendency towards natural desires and enjoying life.

Hofstede's framework, however, has also been criticized for oversimplifying culture into discrete dimensions and ignoring within-country cultural heterogeneity (Kirkman, Lowe, & Gibson, 2006). Empirical studies, though, show most of these concerns are largely debatable (Jones, 2007).

The cultural distance model (Ahn & McKercher, 2015) has been used by researchers in the tourism domain, such as Songshan and John (2019), to study tourists' satisfaction and perceptions of service quality and demonstrated the negative effects of cultural characteristics such as power distance and long term orientation, and another study (Kim & Aggarwal, 2016) found that visiting customers from countries with greater power distance than the destination, exert a sense of superiority to service providers and have high service quality expectations. This highlights issues with cultural distance between the host and tourists' country of residence, a known property influencing satisfaction based on culture gap (Schuckert et al., 2015).

2.3. Purchasing power as a factor that shapes experience

Studies addressing tourist purchasing behavior highlight that the Gross Domestic Product (GDP) per capita of the tourists' country of origin has a significant impact on their satisfaction (Wong & Law, 2003) and expenditure (Sharma, Woodward, & Grillini, 2020), while the GDP of the destination impacts the number of tourist arrivals (Manosuthi, Lee, & Han, 2020).

Evidence suggests that tourists tend to provide more favourable ratings when the level of economic state of the destination is similar to that of their country of origin and reviews of visitors coming from wealthier countries are lower in ratings when on business compared to leisure trips, which is likely related to their higher expectations regarding that aspect of life (Radojevic, Stanisic, Stanic, & Davidson, 2018). This supports evidence that purchasing power is linked with satisfaction (Sharma et al., 2020) and a greater need to portray status through consumption (Dubois & Duquesne, 1993). Given this evidence, it was deemed necessary to examine the relationship of purchasing power with satisfaction.

One way to compare countries regarding spending potential is through purchasing power parity (PPP), used to compare the economic capacity of each country to purchase the same goods and services (Fischer & Lipovska, 2018) and can be measured by the Big Mac Index or GDP per capita. The financial state of a country and the GDP per capita have been used as key indicators for the comparison of national development levels (Gilboa & Mitchell, 2020). Moreover, the GDP dramatically affects quality of life in countries with low human development index score (Dipietro & Anoruo, 2006) and hence could increase expectations. Many researchers argue that tourists from countries with lower purchasing power than that of their visiting destination might be more demanding and hence more likely to evaluate their experience at a destination negatively (Christodoulou et al., 2020).

In light of the above literature and identified gaps, the paper is guided by these research questions:

(i) What topics are discussed in eWOM, by tourists who visited hotels at a holiday destination and how are these associated with satisfaction and dissatisfaction with their experience?

(ii) How are the identified topics from tourists' eWOM related with cultural dimensions and purchasing power (GDP per capita) of the tourists' origin country?

(iii) How can the extracted information be utilised to support message design in DCM?

3. Methodology

The research questions are addressed using a combination of currently evolving big data/social media analytics techniques including *sentiment analysis* for the evaluation of customers' satisfaction through reviews polarity assessment, *topic modelling* for the identification of the main themes discussed in eWOM, and *decision trees* for patterns identification through rules extraction.

Overall, the methodology is composed of three phases, each addressing the following activities: (i) data collection, integration, and pre-processing; (ii) sentiment analysis, topic modelling, data imbalance treatment, feature selection, training and optimization of DTs; and (iii) extracting patterns from DTs and interpreting the results for DCM support. These phases are combined, as depicted in Fig. 1, they are detailed below and implemented within the case study in Section 4.

3.1. Phase 1: data collection, integration and pre-processing

The first phase begins with **collecting** the tourists' eWOM text (i.e., hotel reviews) and **integrating** it with information on the economic and culture conditions of eWOM's authors country. The **data pre-processing** step then involves missing data elimination and data cleansing such as emoticons encoding, spelling corrections, abbreviations expansion, and jargon handling.

3.2. Phase 2: sentiment analysis, topic modelling and construction of optimized DT

Sentiment analysis is performed on reviews to classify them into positive and negative for two distinct purposes. First, sentiment is utilized as topic covariate (i.e., metadata in STM topic modelling) to generate topic-sentiment associations, which in turn are used for the naming of learned topic models' themes. Second, sentiment is used as the class label during the construction of the DTs.

Topic modelling is an unsupervised data mining technique used to identify the main topics mentioned in unstructured data such as online reviews (Korfiatis, Stamolampros, Kourouthanassis, & Sagiadinos, 2019). STM is employed in this study, due to its ability to incorporate metadata, such as sentiment of each review, in the topic model, through covariance (Roberts et al., 2014). The output of topic modelling consists of two lists of topics (one for each hotel category) along with the sentiment polarity of each topic. Alternative supervised techniques (such as Naïve Bayes, Support Vector Machine) to associate topics with reviews are limited because of their dependence on prespecified classes (labels) that are based on a-priori domain knowledge. Such methods lead to good results but require vast amount of labelled data, which is scarce.

Each review is subsequently associated with a topic distribution (theta values from STM model) denoting the percentage of each topic discussed in each review. Topics constitute the features based on which the DTs are trained. The study adopts a binary DT approach associating endogenous eWOM information (topics discussed and sentiment) with exogenous information: culture and economic conditions of authors' country of origin. Tourists' satisfaction is measured as positive or negative sentiment and eWOM's cultural and economic dimensions measured as higher or lower to the touristic destination.

Decision Trees (DTs) refer to a popular classification method that inductively learn their structure by partitioning recursively the training dataset using a series of splits, each performed on the most prevalent variable identified using information gain or Gini impurity index metrics. Splitting variables are used to define the structure of the DT represented by nodes on the tree. Each node splits the dataset into subsets through the rules defined on its branches.. From the various available algorithms, this work employed the CART algorithm to build DTs due to its popularity and performance (Razi & Athappilly, 2005).

In contrast to black-box methods, DTs provide reasoning visibility and they are ideal when explaining the logic behind their predictions (Du, Liu, & Hu, 2019). Hence, DTs explain what inputs are the best predictors of an output, and can indicate the statistical significance and strength of the relationship between input and output variables. Additionally, they have good performance with data where features are individually meaningful (topics refer to themes discussed in reviews hence such features make sense to decision makers) in contrast to applications such as image or speech recognition for which deep learning models are more appropriate (Lundberg et al., 2020). The proposed method combines multiple DT, optimised with feature selection similar

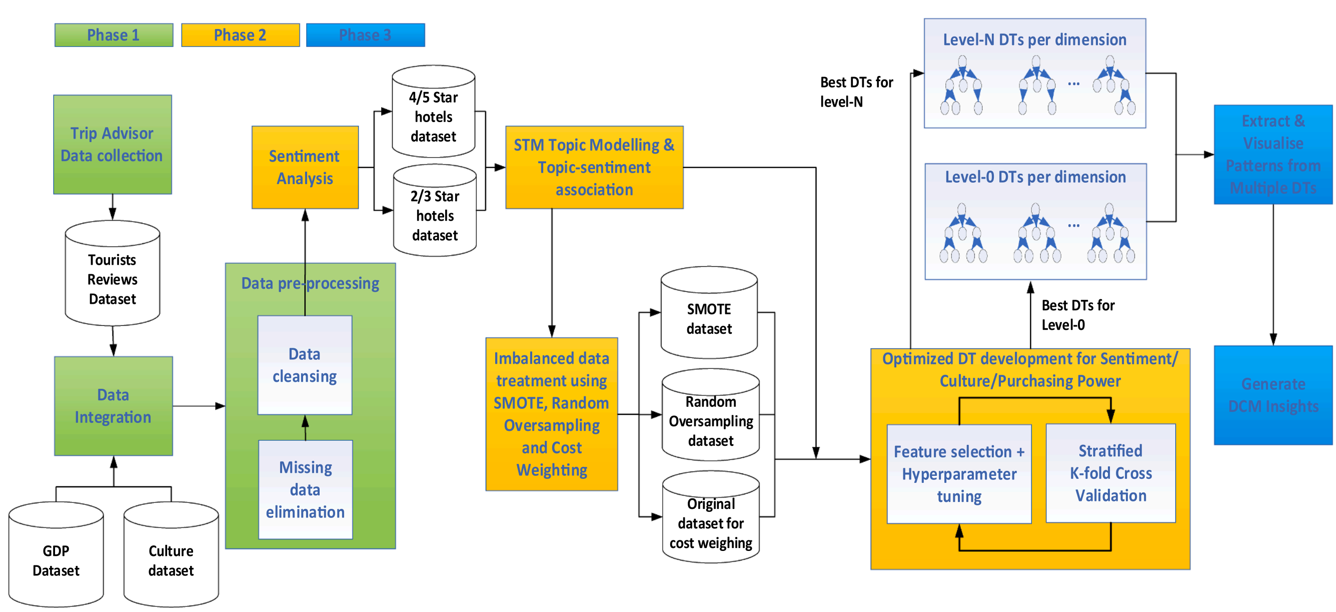

Fig. 1. Overall methodology pipeline.

to (Zhou, Zhang, Zhou, Guo, & Ma, 2021) and imbalance data treatment prior to training, to enhance their classification performance. The optimised DTs are trained using meaningful features (discussion topics) that emerge from topic modelling, hence, makes this combination highly interpretable.

The main drawbacks of DT include creation of over-complex trees that could sometimes overfit and could be biased towards the majority class when the dataset used to train the model is imbalanced (e.g., the number of positive cases is much more than the negative ones). This means it can predict the majority (i.e., positive) class well, but it performs badly on the minority (i.e., negative) class. The three main methods for **tackling imbalanced data**, and employed here, are: (a) cost-sensitive methods, (b) ensembles, and (c) sampling strategies (Gómez, Hernández-Callejo, Martínez, & Sánchez-Esguevillas, 2019). Cost-sensitive methods assign a higher cost to the misclassification of minority cases. Ensemble-based techniques can improve the performance of a classifier by combining several base or weak classifiers (López, Fernández, García, Palade, & Herrera, 2013) and are popular for dealing with the class imbalance problem; however, they suffer from model interpretability due to their black-box nature. Alternatively, distribution balancing can be done by modifying the class distribution of the training data using either under-sampling or oversampling. Oversampling enables the creation of a (more) balanced sample by drawing subsamples from the original dataset and in particular by duplicating samples of the minority class; this is most commonly implemented via Random Oversampling or Synthetic Minority Oversampling Technique (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

Feature selection is an important step in optimising the DT performance. It addresses the process of selecting the variables (features) from the dataset that have the highest impact on the target variable and is proved beneficial in improving DT performance (Wang & Li, 2008). Several techniques exist for feature selection. Heuristics approaches such as simulation annealing (Chopard & Tomassini, 2018) or tabu search (Laguna, 2018) are usually good strategies for near optimal solutions. The three main categories of heuristic feature selection methods are filter methods, wrapper methods, and embedded methods. The filter approaches, are independent of the classification algorithm and select features based on statistical (e.g., chi-square), distance (e.g., relief), or entropy (e.g., symmetrical uncertainty) aspects of the features. Wrapper approaches run a classifier to evaluate feature subsets (Kohavi & John, 1997) while the embedded methods also use a classifier to perform feature selection as part of their learning procedure. In the case study that follows several heuristic methods are utilised.

To evaluate the performance of the DT, several metrics are available. Since this is a binary classification problem using imbalanced data, the main performance metrics are: Area Under the receiver operating characteristic Curve (AUC), sensitivity (recall), and specificity. Given the appropriateness of AUC for such problems (Zhao et al., 2020) it was used during the DT **hyperparameter tuning** process. To obtain the best DT, an exhaustive hyperparameter tuning approach was employed using a grid search method since this approach was deemed safer. The main parameters utilized during tuning were the DT maximum depth, the alpha value (DT cost complexity), and minimum samples per leaf node, and in the case of cost sensitive classification the weights associated with each class (Mantovani, Horvath, Cerri, Vanschoren, & De Carvalho, 2017).

Different DTs were constructed to model the effect of topics discussed in eWOM with each dimension of analysis (sentiment, culture, GDP). The sentiment DT analysis aims to examine which eWOM's topics are associated with increased or decreased satisfaction and in what class of satisfiers (according to the 3-factor theory), these topics belonged to. The DTs for the cultural dimensions and economic conditions aim to identify the topics associated with high and low cultural distance and purchasing power (measured as a difference between host and origin country GDP per capita), since in these cases the effects on satisfaction are more prevalent (Songshan & John, 2019).

The **DT construction process** initiates at the highest level (level 0 in Fig. 1) with the DTs utilising all the features (topics) that emerged after initial feature selection. This is followed by the construction of subsequent level's DT using remaining features not utilised by the DT in the previous level of analysis. The process is repeated until either all features are used, or the AUC performance of the DT generated (after hyperparameter tuning) is not satisfactory (below 75%, thus no further feature exploration is performed). In this way several sets of optimised DTs are developed for each target variable associating as many topics with each variable without jeopardising DT performance.

3.3. Phase 3: extracting patterns from DTs and interpreting the results for DCM support

Pattern extraction is performed by combining the insights from all DTs for each dimension of analysis. The Shapley Additive explanation (Tree-SHAP) model interpretation approach (Lundberg & Lee, 2016) is utilised in this phase due to the numerous DT per dimension of analysis, to summarise the impact of each topic and enable the visualisation of multiple DTs rationales. SHAP is a model interpretation framework based on game theory. Tree-based SHAP can utilise multiple trees and explain the global and local behaviour of these in an intuitive way by computing the contribution of each feature to the target prediction produced by the model in an additive way. Unlike feature selection methods, SHAP can identify whether the contribution of each input feature is positive or negative. SHAP values are generalisations of Shapley values where feature contributions on the model output are calculated based on their marginal contribution (Shapley, 1953). Each tuple in the dataset gets its SHAP value that helps interpret the model's predictions locally or globally (in our case multiple DTs for each target) by summing contributions from their individual feature values (topic thetas). In game theoretic terms, model predictions can be explained by assuming that each feature (topic theta) value is a 'player' in a coalition game (high versus low thetas translated into intense versus insignificant topic discussion in a review) and model's prediction is the pay-out. The Shapley values explain how to fairly distribute the pay-out among feature values. In the proposed method, SHAP values (additive feature attribution) are extracted from all DTs (at all levels) in each dimension of analysis and for all features associated with these DTs. Subsequently, these are analysed holistically, in accordance with Lundberg et al. (2020), to produce the global behaviour in bar-style visualisations that show feature contribution per target variable. SHAP values from the tree explanations are collated in a table associating topics with each dimension of analysis to provide aggregate insights for DCM support.

This white-box nature of our method lends itself to pattern extraction and domain knowledge validation as demonstrated in subsequent sections, in contrast to many black box models such as deep neural nets, which despite their powerful predictions, lack explainability and have limited interpretability, with severe consequences in various applications (medicine, law, policy, energy, finance) (Barredo Arrieta et al., 2020). An example attempt in the tourism domain (Chang, Ku, & Chen, 2020) to classify managerial responses to tourists eWOM using deep learning natural language processing led to mediocre results. Recent studies also show that such models are vulnerable to natural noise and data errors, while they rely heavily on massive labelled datasets to learn useful representations (Shan, Xu, Yang, Jia, & Xiang, 2020), which in turn leads to learning uninterpretable features (Gridach, 2020). Another drawback of deep learning models is that despite the fact that they require large volume of corpus data, it is hard to optimise their model parameters (Alami, Meknassi, & En-nahnahi, 2019). Deep learning was not a feasible option in our case since: 1) our data was not labelled neither too large 2) the use of secondary annotated data based on a different eWOM classification task could bias the analysis and constrain the findings to labels specified in the secondary data.

4. Case study applying the methodology

The methodology is applied on eWOM data, written in English, by tourists in Cyprus who stayed in (reviewed) hotels between 2014 and 2019. The data were retrieved using location filtering criteria in TripAdvisor console and was automatically extracted using a scrapper. TripAdvisor was chosen for three reasons: (i) it is one of the most popular sources of hotel ratings on the Internet with millions of registered users, reviews, and hotels across the globe (O'Connor, 2008); (ii) it is one of the most widely used platforms among the scholarly community (e.g., Yoo & Gretzel, 2008), and (iii) it facilitates filtering of properties and hence suited the goals of this work.

Due to the inherent differences (in service, infrastructure, etc.) among the two main categories of hotels, namely 2/3-star and 4/5-star hotels, the reviews were split into two datasets, corresponding to each hotel category and processed separately. Each dataset was used to develop a topic model that referred to the main themes discussed in the eWOM corpus to build DTs for each class of hotels and dimension of analysis (sentiment, culture, purchasing power).

4.1. Phase 1: data collection, integration and pre-processing

Information from eWOM related to the generator, the hotel, and timing of visit and review, and included the reviewer's nationality based on their self-declared location, their past eWOM contributions, helpfulness vote, hotel rating, hotels star rating, and actual review text. Nationality should be considered only as a proxy of tourists' cultural scores since the location of residence in not always mapped to the country's culture score due to foreigners working in different countries. Given the lack of information on this aspect, and recent evidence on the homogeneity of the scores, even with diverse groups of the same nation (Mazanec, Crotts, Gursoy, & Lu, 2015) we consider this only a small caveat and we do not expect to significantly affect the results. Based on this location and the time of review generation, the following national-level variables were added to the dataset: level of economic development (as measured by nominal GDP per capita in US$ at the time of the review) and cultural characteristics (as measured by scores on the six cultural dimensions) for both Cyprus (i.e., the destination) and the reviewer's home country. By including Hofstede's culture scores at the national level, we consider culture as a collective, and not an individual phenomenon. The underlying assumption with the chosen approach is that Hofstede's scores (at national level) represent a proxy of the sub-population of a country's citizens that take their holidays in Cyprus.

Since this study focuses on cultural distance between the visitor's country of residence and the host country (Songshan & John, 2019), these differences were calculated utilising counties' scores from Hofstede website,[4] (cultural dimension values expressed in a scale from 0 to 100 for each), based on year of review. The values of purchasing power are also expressed at the national level (Christodoulou et al., 2020). To estimate each country's purchasing power, the GDP per capita index was used with data from the World Monetary Fund. The metric referred to GDP performance per country on a yearly basis. The variable is expressed in US dollars and was standardized in a scale from 0 to 100 across all reviews.

During data preparation, the endogenous part of eWOM (textual information) had to be pre-processed further to eliminate irrelevant information through stop-word removal, stemming, tokenization, punctuation removal, custom words removal, numbers removal, and converting all text to lowercase. Initially, common stopwords were considered and gradually with the refinement of the model, additional stopwords that were irrelevant to our goal were added to the list of custom stopwords, similarly to (Gregoriades & Pampaka, 2020). These included names of hotels, cities, and resorts. Tokenization, and in particular the use of phrases composed of *n*-words (i.e., *n*-grams) was applied to transform the reviews into a sequence of tokens. The Stemming process (i.e., the conversion of words to their root form), was not implemented due to reported questionable benefits (Schofield & Mimno, 2016) especially for the purposes of this analysis.

Finally, tourists' eWOM text, time dependent GDP per capita, and numerical values of culture's dimensions were integrated to form a collated dataset used for DT training.

The data also underwent rescaling and missing data elimination (i.e., culture values not available in Hofstede database). Reviews from local tourists were also eliminated, since the cultural distance theory does not apply to them since the distance is zero.

4.2. Phase 2: sentiment analysis, topic modelling and construction of optimized DT

This phase focuses on the evaluation of reviews' sentiment and the development of the topic models for each class of hotels. The training of the binary DTs is optimized using hyperparameter tuning, imbalance data treatment, and target variable discretization.

4.2.1. Assessing satisfaction through sentiment analysis

Sentiment analysis was used to assess eWOM's satisfaction and as metadata during topic modelling. Two trained sentiment models were used for this task in the same manner as Gregoriades and Pampaka (2020): Textblob (based on Naive Bayes) and Vader, which are popular sentiment classifiers with good performance used in similar research (Sadiq et al., 2021). To improve our confidence in the sentiment analysis, both models were used in combination through a python script that averaged their results. The process was repeated for all reviews in the dataset, and the polarity of each review was added as a new variable in the dataset.

An issue that negatively affects classification models is that of data imbalance that refers to datasets that have more cases of one type over the others. In our case, the distribution of the reviews over sentiment classification was heavily imbalanced: 5% negative, 10% neutral, and 85% positive. Hence, to reduce this imbalance, sentiment polarity was converted to a binary variable by collapsing the neutral and negative sentiment categories of reviews. This resulted in data distribution which is still imbalanced (15% negative-85% positive), an issue addressed in Section 4.2.4.

An additional issue relating to data imbalance in this case related to the country of origin/residence of the reviewers: the vast majority regarded tourists from the UK. To address this issue prior to modelling and to minimize the potential related biases, a random sample from the UK group of equal size as the maximum number of reviews from other tourists' groups was selected. This yielded a dataset of 42 K reviews.

4.2.2. Topic modelling

STM topic modelling was performed separately for each of the two datasets to obtain topics that best relate to each hotel category since different categories are associated with different topics (e.g., spa topics are not relevant to 2/3-star hotels). In each STM model, each review is associated with a distribution of a finite set of topics; topics are distributions of words in the corpus of all reviews. The probability distribution of topics per review denotes the probability of each topic discussed in a review and the sum of all topics' probabilities in each review totalled to 1. An important task in topic modelling is the selection of the optimum number of topics (k) to identify in a corpus. There are several approaches to this problem. In this work, different models were built using different number of k (10–100): these were evaluated in terms of exclusivity and coherence (Roberts et al., 2014) to find the optimum value for k as depicted in Fig. 2. Coherence denotes whether words under the same topic make sense when they are put together while exclusivity refers to how exclusive a word is to a topic (i.e. not being common in other topics). Based on the results and relevant

[4] https://www.hofstede-insights.com/product/compare-countries/.

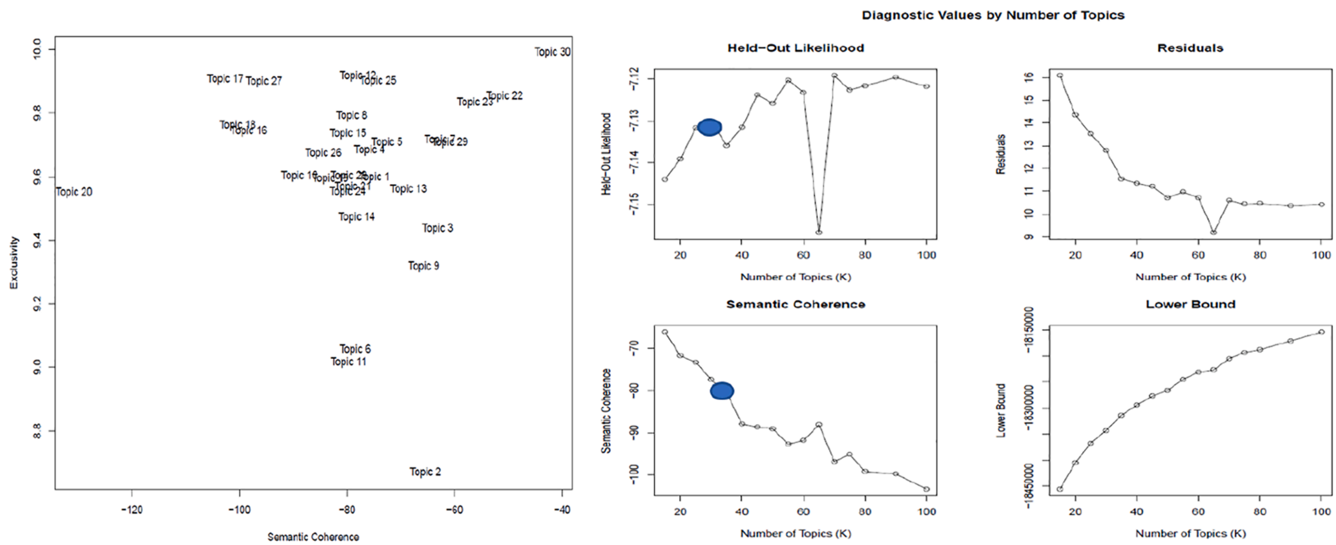

Fig. 2. Selection of the optimum number of topics k (right) based on semantic choherence and held out likelihood and residuals optimisation. Topics coherence and exclusivity after developong the STM with k set to 30 for the 4/5 star hotels dataset (left).

recommendations (Roberts et al., 2014), we consider 30 topics for the 4/5 hotels and 35 topics for the 2/3 star hotels to capture topic diversity without impeding topic interpretability. These analyses were conducted with the STM package (Roberts, Stewart, & Tingley, 2019) in R.

The identification of words best linked together forming a distinctive component relating to documents (reviews) in a corpus is realised through a statistical algorithm and hence it is entirely reproducible; the interpretation and labelling of topics, though, is more subjective and clearly relies on the analysts' expertise and prior knowledge. A popular approach for labelling topics, also used herein, is to first consider words highly associated with each topic and secondly to inspect the most prevalent reviews related to that topic. For the first task, extracted topics were inspected to make the necessary connections between the main words associated with the topic that emerged from different metrics such as highest probability, Lift, and FREX (Roberts et al., 2019). Lift weights words by giving higher weight to words that appear less frequently in other topics and FREX weights words by their overall frequency and how exclusive they are to the topic.

Topics naming was motivated by prevalent hotel service quality factors in hospitality literature (Banerjee & Chua, 2016) such as convenient location, service quality, reputation, and friendliness of staff and factor groupings such as tangibles (e.g., equipment), reliability (e.g., punctuality), responsiveness (e.g., prompt service), assurance (e.g., politeness), and empathy (e.g., personal attention). To assist the interpretation of topics, it was essential to identify the polarity of each topic based on its association to satisfaction (sentiment). To that end, the STM's metadata capability was used with sentiment set as prevalence variable (metadata) to assess the association of each topic with the outcome (sentiment) using LASSO based estimates[5]. For topics that were not included in the LASSO output a logistic regression model was generated per topic with the reviews' sentiment as the dependent variable and the topics' theta values from STM models as independent variable. The evaluated topics' polarity helped to refine the naming of topics to have either positive or negative connotations, to later enable the interpretation of the DTs. The main topics that emerged from the analysis are depicted in Table 1.

The trained STM model was used to extract the associations between topics and reviews. Each topic was appended as new variable in the dataset similarly to one-vs-rest binarization strategy (Yan, Zhang, Lin, Yang, & Luo, 2020) but instead of zero values for not focal variables, topic-variables were associated to the theta values (representing the distribution of topics over reviews) of each topic per review that refer to the probability that a topic is associated with each review

4.2.3. DT target variable discretization

Prior to DT learning, the target variables were discretised into 2 states (higher/lower culture or economic state to country of residence and destination). Binary classification aims at finding a mapping from an input vector space to a discrete decision space that has only two states and was adopted due to its suitability to our problem. The sentiment target was classified as positive and negative. The discretisation for the cultural dimension targets was performed using the host country's culture and GDP per capita scores (Cyprus) at the time of each review as the threshold values to calculate the cultural distance and purchasing power difference from the visitor's country, as either "high" (if higher than Cyprus) or "low". This discretisation approach would involve removing cases with the same score as Cyprus, however there were no cases like that in our dataset. Target variables are expressed in binary form because this discretisation is giving the best DT performance and is easier to interpret. The discretisation of cultures' target variables was validated against different threshold values as demonstrated in subsequent section to verify that the split criterion was associated to a value that highlighted the cultural distance between tourists and host country. Finally, dataset outliers relating to topics associated with extremely high or low thetas in reviews were eliminated since this yielded the best DT performance.

4.2.4. Imbalance data treatment

Due to the imbalance nature of the datasets that emerged after preprocessing, it was necessary to balance these to minimise model overfitting during DT training. SMOTE, random oversampling, and cost sensitive approaches, were tested, with the cost sensitive method producing the best results (according to AUC). SMOTE yielded DTs that overfitted the data, while random oversampling did not achieve a satisfactory AUC.

4.2.5. DT training and performance optimization

Prior to training the DTs two activities were performed to optimise classification performance: feature selection (Li et al., 2018) and hyperparameter tuning.

In feature selection, all three categories of heuristic techniques were used for feature selection as described in the methodology. Hence,

[5] Using the R package STM: https://cran.r-project.org/web/packages/stm/stm.pdf.

Table 1
Naming of topics derived from the two STM models for four/five (right) and two/three star hotels (left).

Defined Topics for 2/3-star hotels based on STM model	Evaluated Topic Polarity	Defined Topics for 4/5-star hotels based on STM model	Evaluated Topic Polarity
V1_clubbing_holidays (Noise)	Negative	V1_revisited_refurnished	Positive
V2_renovation	Negative	V2_bar_food_cats	Negative
V3_good_rooms	Positive	V3_nice_pool_drinks	Positive
V4_dirty_bathroom	Negative	V4_spa_gym_massage	Positive
V5_dirty-room-unprofessional_staff	Negative	V5_staff_always_willing_to_help	Positive
V6_helpfull_staff	Positive	V6_rude_staff	Negative
V7_other_guests	Negative	V7_excellent_service	Positive
V8_limited_options_all_inclusive	Negative	V8_best_stay_ever	Positive
V9_good_location	Positive	V9_nice_room	Positive
V10_amazing_staff	Positive	V10_extra_costs_for_ameneties (WIFI, coffee machines)	Negative
V11_not_well_equipped_room(e.g. fridge)	Negative	V11_extra_charges	Negative
V12_good_mobility_options(bus, walk)	Positive	V12_amazing_dinners (buffet, a la carte)	Positive
V13_comfortable_room	Positive	V13_pool_area_issues (kids, noise)	Negative
V14_poor_entertaiment_and_food	Negative	V14_good_transportation_options	Positive
V15_limited_breakfast_options	Negative	V15_located_close_to_beach	Positive
V16_pool_area	Positive	V16_great_wedding_venue	Positive
V17_value_for_money	Positive	V17_birthday/anniversaries	Positive
V18_close_to_beach_with_pool	Positive	V18_perfect_location(walk around)	Positive
V19_basic_accomotation	Negative	V19_disagree_with_hotel_reviews	Positive
V20_accept_late_arrivals	Positive	V20_luxurius/quality -amenities	Positive
V21_good_dinner	Positive	V21_great_team_of_staff	Positive
V22_low_quality_food_and_drinks	Negative	V22_room_with_seaview	Positive
V23_smelly_room	Negative	V23_lovely_staff	Positive
V24_good_bar_service	Positive	V24_dirty_room_beddings_ shower	Negative
V25_ideal_for_summer_holidays (pool_area)	Positive	V25_missing_services	Negative
V26_bad_customer_service	Negative	V26_great_entertaiment (bingo, darts)	Positive
V27_old_room	Negative	V27_great_for_families	Positive
V28_good_entertaiment	Positive	V28_checkin/front_desk_issues	Negative
V29_friendly_staff	Positive	V29_beach_hotel_with_convenient_location	Positive
V30_will_revisit	Positive	V30_exceptional_staff-food-pool	Positive
V31_basic_apartment_no_luxuries	Negative		
V32_the_best_experiance	Positive		
V33_close_to_beach	Positive		
V34_bad_facilities_maintenance	Negative		
V35_clean_good_service	Positive		

adaptive boosting, random forest, and logistic regression were used for wrapper methods, ANOVA (F value) for filtered methods, and LASSO for embedded methods. The random forest method resulted in DTs with slightly better AUC performance when combined with hyperparameter optimisation, as illustrated in Fig. 3, and hence was adopted.

Hyperparameter tuning is performed to identify the optimum configuration of hyperparameters (DT depth, alpha, samples per leaf, and in the case of cost sensitive classification the class weights) for near optimum model performance using AUC as metric. Hyperparameter tuning was performed in both scenarios (2/3 and 4/5-star hotels) for each target variable (sentiment, cultural dimension, purchasing power). Given the multiple DT training per dimension of analysis that was adopted, as expected, when DTs were trained using all topics that emerged after feature selection, the AUC was higher (the top level of analysis). As topics were removed from the initial dataset (features used to develop the top-level DT), the AUC of subsequent levels' DTs dropped since these models were developed using less influential features.

The stratified k-fold cross-validation approach was used to evaluate the DTs that emerged from hyperparameter tuning and imbalance treatment. This approach splits the dataset in k-folds, each having a percentage of the target variable class and then using combinations of folds for training and remaining folds for testing. This process increased our confidence in the generated DTs. Fig. 4 illustrates the k-fold validation performed for one DT. The higher the area under the ROC curve the better the model's performance. The best DTs were obtained with the cost sensitive approach (imbalance treatment) using random forest

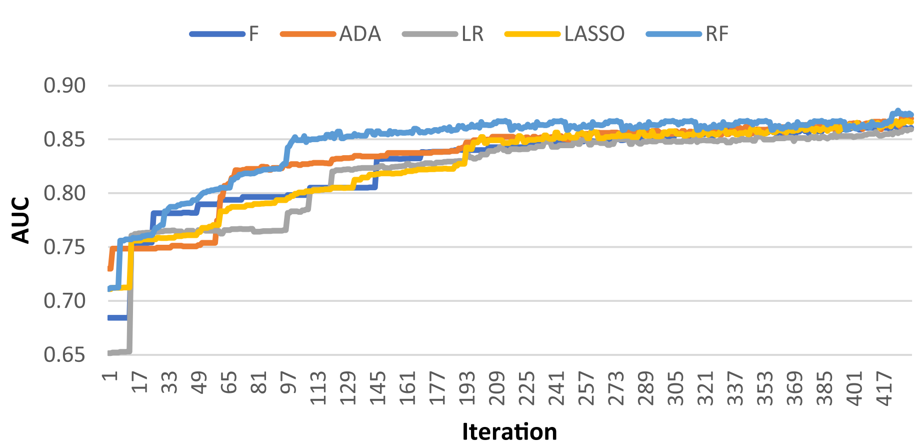

Fig. 3. AUC DT performance per feature selection method. On the x axis is the number of iterations performed until hyperparameter tuning converged to near optimum configuration of the DT that achieved the best performance.

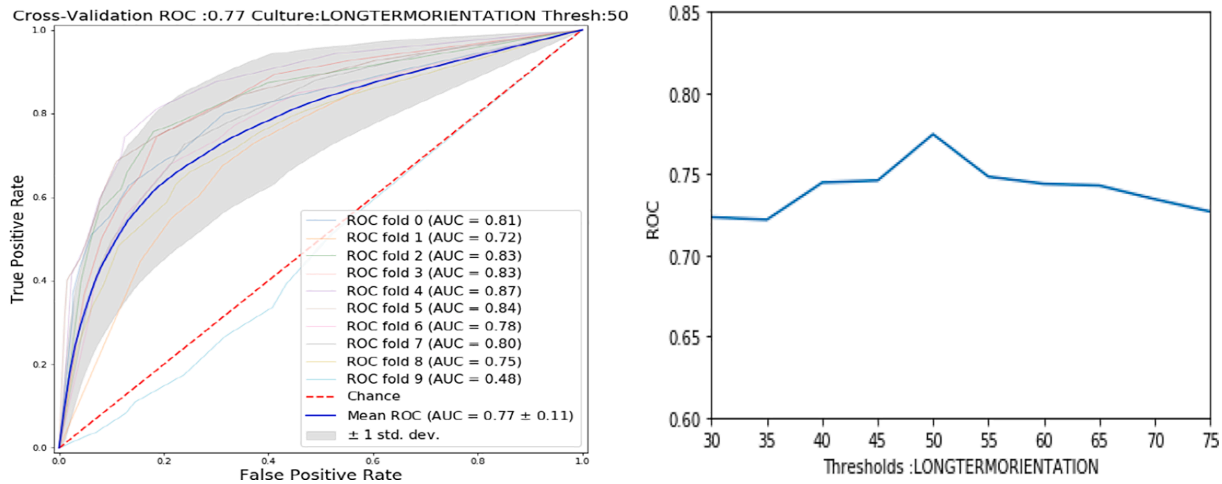

Fig. 4. 10-fold cross validation and verification of threshold value for the Long-term Orientation trait.

feature selection with AUC for the sentiment DT (2/3-star) being 88.9%, while the DT generated from the SMOTE dataset was inferior (80% AUC) due to borderline cases regeneration.

The validation of cultures' target variables discretisation threhsold was perfromed using startified k-fold cross validation and varying target cut-off values. Fig. 4 depicts an evaltuation of *Long-term Orientation* DT against different thresholds and with best hyperparameters and feature selection to verify that the optimum AUC was achieved when the threhold value was set near the host country's cultural score and hence confirming the theory of cultural distance (Songshan & John, 2019). This proved that the DT performance is best when the target variable's binarization is made with threshold value being the host country's score.

4.3. Phase 3: extracting patterns from DTs and interpreting the results for DCM support

The learned DTs were used to identify patterns in the data expressed as rules, combining variables indicated as nodes on the DTs. Nodes also provide information concerning target's class and are color-coded accordingly. In the case of satisfaction analysis, the target variable had positive or negative states and in the case of culture traits and purchasing power the target variables had high or low state indicating higher or lower cultural/purchasing power score of tourists' country of residence over the host country. The emerged rules are used to find topics associations with different cultural dimensions and purchasing power to assist in specifying the content of messages during DCM. Such information can provide useful marketing insights to management to evaluate various aspects of their services provisions (e.g., their best and worst performance), and accordingly tailor their marketing campaigns considering their qualities and also the particular groups they can satisfy more effectively.

The sentiment classification task for each hotel group dataset and associated DTs revealed that the predictor variables (Topics) were able to predict group membership (reviews' sentiment polarity) better than

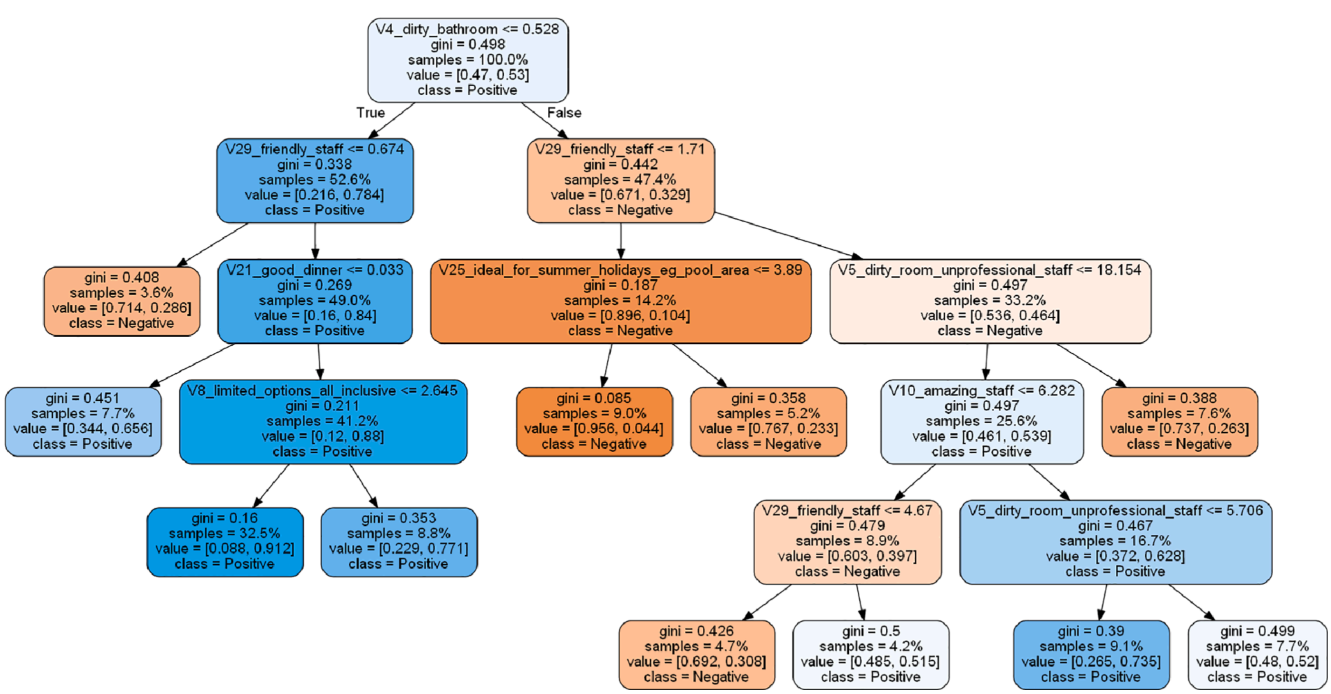

Fig. 5. Learned Decision Tree for 2/3-star hotels. Sentiment predictions colour coded with negative sentiment shown in shades of orange and positive in shades of blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chance (50/50). For the 2/3 star scenario the overall DT's AUC for sentiment classification was 88.9% for top level topics analysis and 83% for level 1, while for 4/5 star hotels the AUC was 90% and 85%, respectively. The structure of the sentiment DT (2/3-star hotels) for the top-level topics is depicted in Fig. 5. Each node represents the state of each topic expressed as a percentage with zero indicating no discussions about a topic in reviews and 100 extensive discussions. The values in each tree node denote the number of cases from the dataset that satisfy a given node criterion (samples %). CART DTs utilise the Gini index or information gain (using entropy/uncertainty) as a splitting criterion, the former calculated as [1 minus the sum of the squared probabilities of each class]. Gini is a measure of the degree of impurity of each node's distribution (when all cases belong to the same class then purity is maximised); lower values indicate a larger difference between negative and positive sentiment reviews. This is also visualised using colour coding on the tree nodes (i.e., the intensity of the colour is associated with impurity). The Gini index is utilised to identify strong patterns in the dataset that are expressed in the form of rules. Based on the figure, there are a few dominating rules that can be derived, each associated with satisfaction and dissatisfaction.

To explain the aggregate effect of all rules that emerged from the DTs in each level of analysis, the SHAP approach (Lundberg & Lee, 2016) was adopted. SHAP values were initially extracted from all DTs (at all levels) in each dimension of analysis and for all features associated with these DTs. Subsequently, these were analysed holistically to assess their impact on the target variables prior to being visualised in a bar chart (Fig. 6). Through this holistic analysis, several DTs are collated in a single table with the dominating rules in each DT competing based on their impact on the target variable. While rules are useful in understanding the rationale of the model, when it comes to providing decision support for DCM message design, it is easier to convert the information about the impact of topics on each dimension of the problem into actionable results, hence the use of tables in this study.

The dominating topics in the top-level sentiment DT depicted in

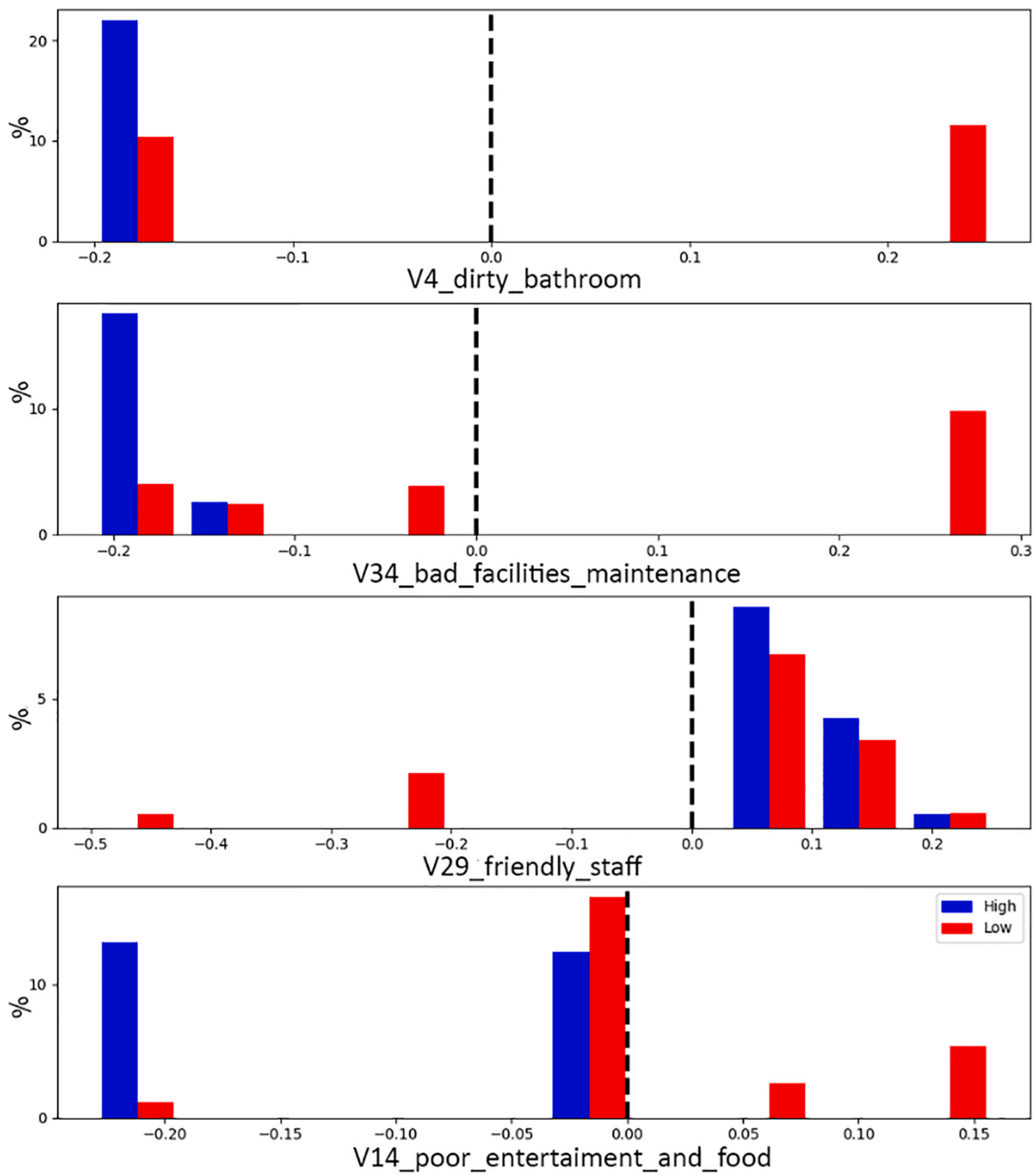

Fig. 6. SHAP summary plot with SHAP values on x-axis for sentiment DT-set for 2/3-star hotels. Blue bars refer to high percentage of a review talking about that topic and red to low topic discussion. Topics are in order of importance measured by the mean absolute Shapley values and reveal how the topic can change the predicted absolute probability of target variable (positively or negatively) by percentage points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5 are the topics associated with room cleanliness, staff professionalism, pool areas, and food. The strongest predictor of dissatisfaction is the combined effect of staff unfriendliness and poor cleanliness, yielding 95% negative reviews (computed based on negative cases over total cases on the leaf node), while friendly staff combined with clean room and good dinner yield 91% positive eWOM. This abides with previous results (Sann, Lai, & Liaw, 2020) stressing the importance of staff performance on satisfaction. An aggregate view of all DTs relating to each dimension using SHAP summary plot revealed the most important topics in all DTs levels per problem dimension/target according to a pre-specified threshold value to focus only on top rated topics. The plots (Figs. 6, 8, 10) depict the topics in ascending order of impact based on their mean absolute Shapley values that reveal how the topic can change the predicted absolute probability of target variable (positively or negatively for sentiment) by percentage points. Each plot depicts on the x-axis the SHAP impact on the target variable with negative values associated with negative state of output variable (i.e., dissatisfaction). The bars' labelling indicate the percentage of topic's discussed in reviews, with red bars relating to low and blue to high (>50% of review talking about that topic). The height of each bar refers to percentage of either high or low topic labels associated with reviews. In Fig. 6, for example, when tourists highly discuss the topic "dirty-bathroom" in reviews, the sentiment of the review is negative, while when this topic is not discussed highly in reviews or discussed infrequently, the sentiment is positive. In the latter case, when the review theme is strongly associated with "staff friendliness", the marginal negative effect of the previous topic ("dirty room") is cancelled out by the strong positive effect of this topic and the end result is a positive review as visualised in the DT of Fig. 5. A vertical dashed line divides the positive from negative associations of topic intensity to target variables. Hence, in Fig. 6, bars appearing on the right side of the vertical line are associated with positive sentiment since in Fig. 6 values on x-axis are positive. The analysis is conducted for all culture dimensions and GDP per capita but only the sentiment and power distance collated SHAP charts are presented analytically, due to space limitations.

These results comply with the 3-factor theory, with topics intensity in eWOM associated asymmetrically with performance of basic satisfaction factors (in the 3-factor theory). These are related to cleanliness and hotel facilities. Topics that occur only in the positive sentiment category could be considered as exciting factors since their absence is less likely to lead to dissatisfaction (e.g., great dinner quality) but their presence has an impact on satisfaction.

Fig. 8 depicts the most influential topics to satisfaction based on the analysis of all DTs associated with the sentiment dimension for 4/5 star hotels. The intensity of eWOM regarding unhygienic settings near bar food (e.g., cats) leads to negative sentiment, while eWOM on staff willingness to help leads to positive sentiment. However, the effect of unhygienic settings is stronger than staff helpfulness alone. Rude staff also has a strong impact on dissatisfaction along with extra charges. The combined effect of topics on satisfaction can be visualised on the DTs with an example top level DT depicted in Fig. 7.

The same approach is followed for each of the cultural dimensions, and purchasing power. Based on the analysis of the DT associated with the Power distance trait for the 2/3 star hotels category, when eWOM topics refer to service-oriented issues such as entertainment, late arrivals, all-inclusive options, and customer-service, the probability that the guests belong to lower power distance class than the host country is increasing as indicated in the SHAP summary plot in Fig. 10. Higher power distance customers are more likely to include in their reviews, cleanliness complaints induced from infrastructure issues (i.e., smelly room). The top-level DT for this trait is depicted in Fig. 9 visualising a subset of the insights from Fig. 10. The intensity of the orange nodes refers to guests having lower power distance scores than the host country. This highlights that when the topics discussed refer to "good entertainment" and marginally about "accepting late arrivals" or the hotel being "close to the beach" then the probability that the guests come from a country of lower power distance than the host country is 98%. This can be expressed as probability P(power distance = lower than host country | good entertainment > 2.8 & Accept late arrivals < 1.8 and close to the beach < 1.4) = 0.98, or simplified to P(PD = Low | Late-Arrivals = Low & Beach-Hotel = No) = 0.98.

A collated view of the insights extracted from all DTs for all dimensions of the problem is depicted in Tables 2 and 3. Based on these tables, a hotelier in Cyprus who aims to target UK tourists must first find how cultural scores of UK relate to Cyprus scores and subsequently identify topics from the table that could positively affect the recipients of his/her message. For instance, if a 4/5-star hotel targets UK tourists, then the cultural and purchasing power distance between UK and Cyprus needs to be identified from Hofstede and the World Monetary Fund. In this scenario, UK's GDP per capita is higher than Cyprus, Power distance is lower, Individualism is higher, Masculinity is higher, Uncertainty avoidance is lower, Long Term Orientation is relatively equal (hence is

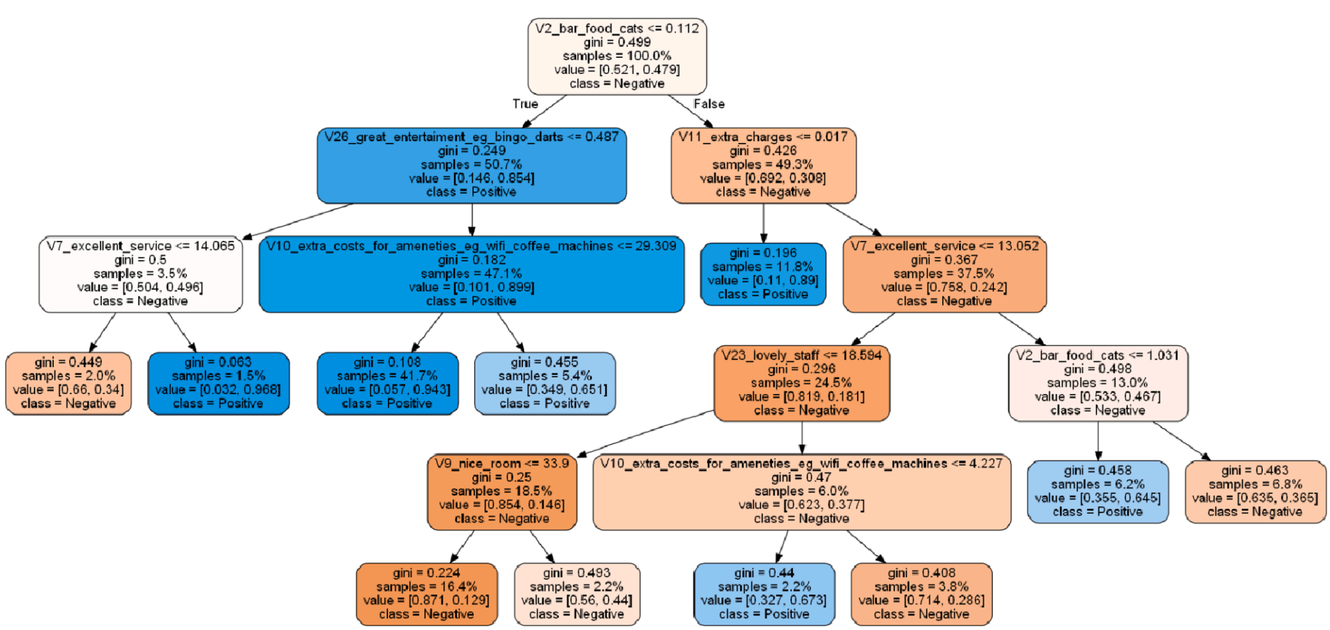

Fig. 7. Sentiment top level DT for 4/5-star hotels.

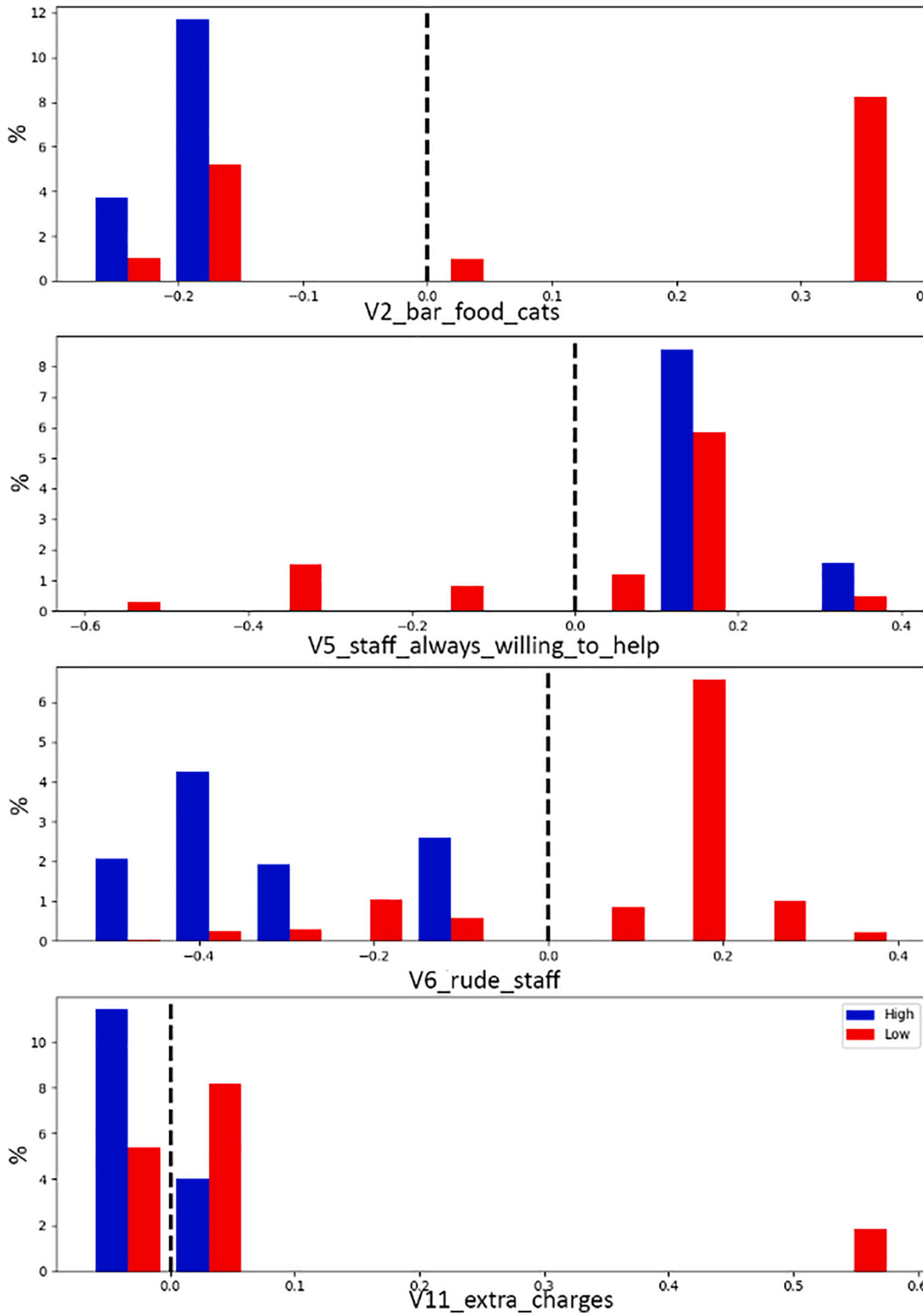

Fig. 8. Sentiment SHAP summary plot for 4/5-star hotels.

not used), and Indulgence is higher. Thus, based on Table 3, the topics that can be used by the DCM team during marketing message design, in order of importance are: nice pool drinks and food, nice pool areas, spa services, sea view rooms, professional staff attitude. Using the same logic for Russian tourists, the message could highlight issues pertaining to: no extra costs for amenities, sea view rooms, spa and gym services, luxurious amenities, and friendly staff attitude. In addition, hoteliers should also consult topics associated with low sentiment, so in the case of 4/5 hotels customers dislike unhygienic practises during bar food serving (e.g., cats around food area) and extra charges for basic amenities such as Wi-Fi, while for the 2/3 star hotels the main complaints concentrate on cleanliness, facilities, entertainment, and bad customer service as shown in Table 2. These insights should also be incorporated in targeted messages by highlighting tourists' likes and negating or inversing identified dislikes.

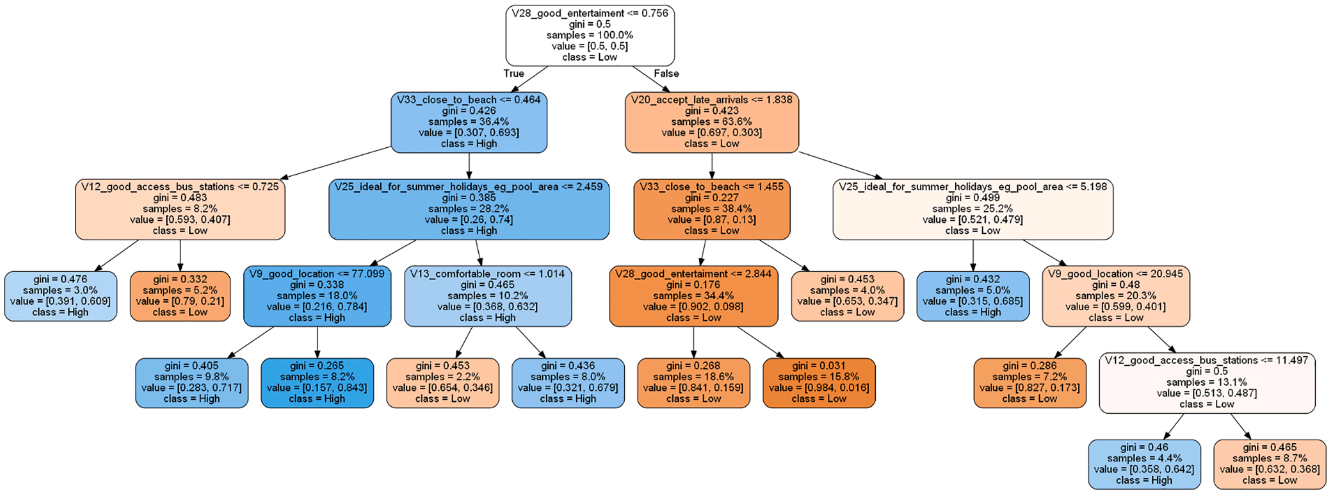

Fig. 9. Power distance top level DT for 2/3-star hotels.

5. Validation against domain knowledge and other white-box classification techniques

To increase our confidence in the results, we first verified them against domain knowledge and subsequently against three white box classification techniques, namely Bayesian Belief Networks (BBN), Naïve Bayes, and Ridge classification.

For the domain knowledge validation, insights for each cultural trait were compared against the literature. Regarding the high individualism trait, the results show that it is highly associated with the topics: nice pool drinks, entertainment, and professional staff, which are all positive topics and hence confirming previous findings (Johnson, Herrmann, & Gustafsson, 2002) that this trait is associated with higher levels of satisfaction especially for hotel services (Frank, Enkawa, & Schvaneveldt, 2015). The high power distance trait was found to be associated with service complaints in contrast to low power distance and hence confirming the findings of Banerjee and Chua (2016) that tourists from countries with higher power distance than the host country are more likely to complain about the level of service. In the case when both guest and host are from cultures characterized by high power distance, the satisfaction decreases further (Huang, Huang, & Wu, 1996). This finding is also supported by our results, with Cyprus having high power distance score (Epaminonda, 2020), when visited by tourists that come from even higher power distance it leads to complaints about room cleanliness and extra charges. People with high uncertainty avoidance tend to avoid risk and this was verified by the strong association of this trait with the topic "acceptance of late arrivals" (Reisinger, 2009). Long-term orientation as a more tradition-based trait is associated with topics relevant to group of people such as family holiday (e.g., pool area), entertainment, and good staff interaction, while short term orientation with individualistic activities such as gym or good location to be able to explore easily the surrounding attractions. The masculinity trait was associated with bar service and drinks while femininity trait with complaints for extra charges. Finally, the indulgence trait that is related with "enjoying life and having fun" (Swierstra & Rip, 2007) was associated with topics relevant to food and drinks.

For the quantitative evaluation of the DTs, the AUC score is used to compare the performance of three popular classification techniques with good interpretability characteristics. For the development of the BBN model, the WEKA data mining tool was used (Witten, Frank, & Hall, 2011), while for the Naïve Bayes and Ridge classification, the scikit-learn machine learning library was used. The BBN type used is the General Bayesian Network (Madden, 2009) where each topic represents a node on the BBN. Prior to BBN structure learning topic variables underwent discretisation using equal intervals and equal frequency options with different number of states. BBN structure learning is performed using search-and-score based techniques similar to Madden (2009) utilising two options: the K2 algorithm with the Bayesian–Dirichlet equivalent uniform (BDeu) scoring and Hill-climbing search with the Minimum Description Length score (Koller & Friedman, 2009). BBN structure learning is combined with the variable discretisation methods to find a combination yielding the best AUC performance. Since there were no missing values in the dataset the BBN's conditional probability tables were computed using WEKA's Simple estimator that uses the relative frequencies of the associated combinations of topics' states in the training data. The best AUC score from these options (structure and discretisation) is used as the performance of the BBN in the table. Similarly, the hyperparameters of Ridge and Naïve Bayes models were optimised in python using grid search. The results of the experiments for the 2/3 and 4/5 star hotels are depicted in Table 4. The DT marginally outperforms the BBN and Ridge classifiers in AUC in the 2/3 star dataset, while being marginally behind Naïve Bayes at two cultural dimensions. In the 4/5 star dataset the DT marginally outperforms most of the classifiers in most target variables while being marginally behind Naïve Bayes at three cultural dimensions. However, despite its good performance in both categories of hotels, Naïve Bayes is based on the assumption that input variables are conditionally independent of each other and hence the interactions among features are not considered when predicting the output, which limits the insights that can be extracted from such models. In DTs, the relationship between features and outcome could be nonlinear with features interactions visualised in a comprehensible way and insights expressed in the form of rules that could incorporate more than one feature.

6. Conclusions and discussion

In e-marketing, the content of the message plays an important role in the formation of a customer's attitude towards the message source. Automated marketing communication enables firms to gain high market share and reputation with less investment (Luo & Donthu, 2006), which can lead to purchase intention and positive eWOM. Irrelevant or redundant messages, however, can cause customer frustration (Wang, Xiong, & Olya, 2020). Several black box algorithms have been used in this endeavour, however, these hide the reasoning behind the algorithm's logic and hence increase the distrust of decision makers who cannot understand the rationale of the predicted/modelled decisions. White box techniques such as DTs provide marketing managers with an interpretable explanation of the algorithms' reasoning, which consequently increases their trustworthiness. An important component that is missing from existing algorithms and supports message design is the

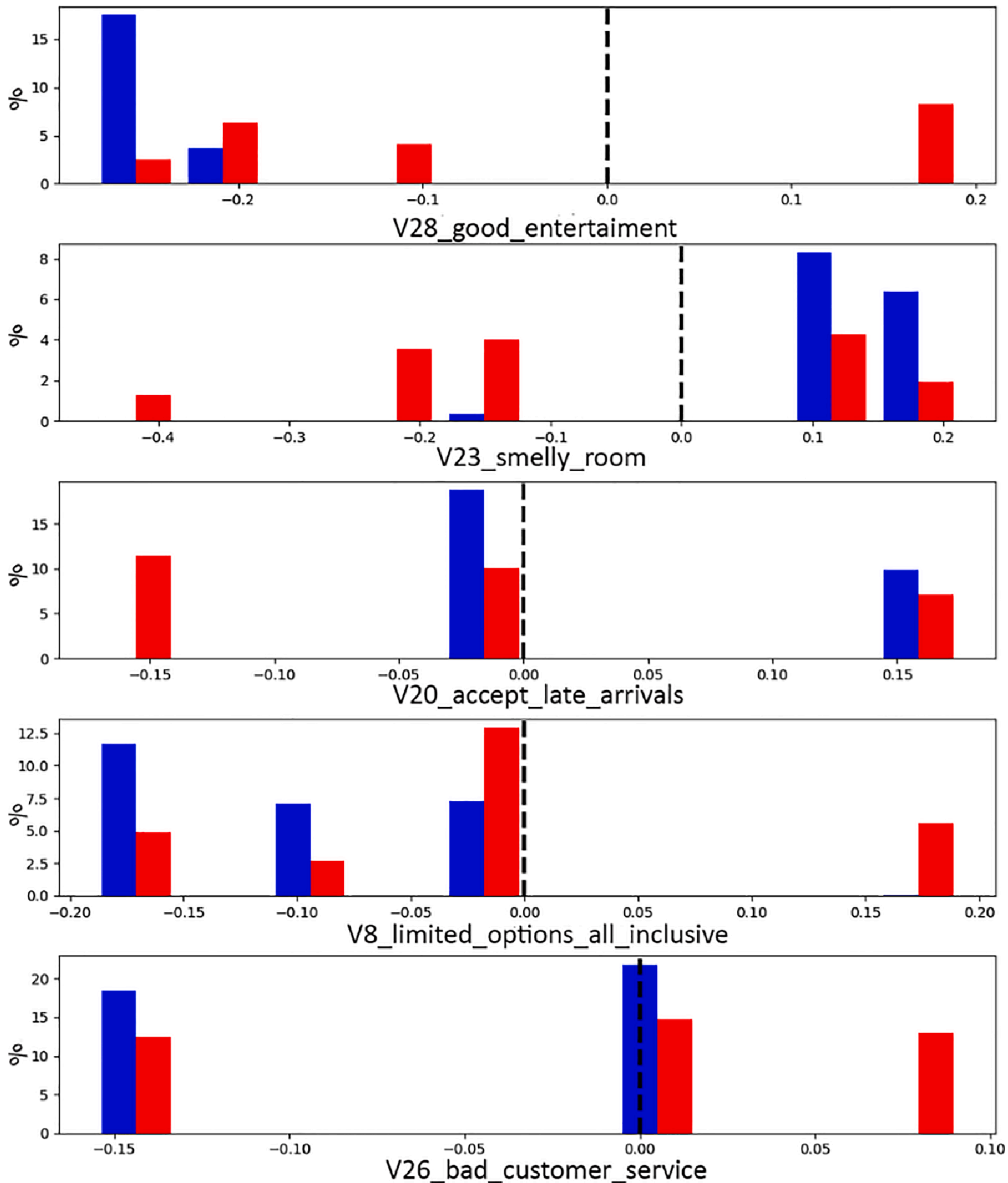

Fig. 10. Power distance SHAP summary plot for 2/3-star hotels.

Table 2
Sentiment associations with topics. Topic importance measured is indicated by the mean absolute Shapley values in parenthesis and reveal how the topic can change the predicted absolute probability of satisfaction (positively or negatively) by percentage points.

	2/3-star hotels	4/5-star hotels
Positive Sentiment	Friendly staff (0.12), will revisit (0.08), good dinner (0.03), amazing staff (0.03), close to the beach (0.02), ideal for summer holidays (0.01)	Staff always willing to help (0.20), great team of staff (0.05)
Negative Sentiment	Dirty Bathroom (0.23), Bad Facilities maintenance (0.21), poor entertainment and food (0.06), bad customer service (0.05), dirty room and unprofessional staff (0.04), limited options of all inclusive (0.02)	Bar food with cats (0.27), rude staff (0.17), extra charges (0.1), extra cost for amenities (0.05)

culture of the message recipient. Existing literature mainly examined the impact of culture on review generation (Kim, Jun, & Kim, 2018; Koh et al., 2010; Purnawirawan, Eisend, De Pelsmacker, & Dens, 2015), but to the best of our knowledge there are no studies identifying what factors (topics) inferred from eWOM are associated with which cultural dimension or purchasing power to inform message design; an aspect that makes this work novel. The proposed method presented identifies patterns among themes discussed by tourists based on their different culture dimensions and purchasing power through the combination of DTs with topic modelling and sentiment analysis to assist DCM in designing relevant messaging that would better attract customers. The method is bottom up and infers patterns from eWOM in contrast to top-down approaches that utilise surveys or aggregate hotel ratings from prespecified categories of user satisfaction attributes scales. Top-down approaches constrain the analysis to the dimensions provided by the data collector/ provider. Other scholars also used a bottom-up approach through topic modelling to identify themes discussed in reviews but not to investigate topics' relationship with culture, purchasing power (i.e., GDP per

Table 3
SHAP values associating topics with cultural and purchasing power. Topic importance measured by mean absolute Shapley values in parenthesis.

	2/3-star hotels	4/5-star hotels
Higher GDP	Ideal for summer holidays (0.15), good entertainment (0.14),	Nice pool drinks (0.17), pool area issues (0.8), missing services (0.03), revisited refurnished (0.01)
Lower GDP	Good location (0.19), accept late arrivals (0.17), comfortable room (0.08), old room (0.08), bad customer service (0.01), clean good service (0.01), poor entertainment (0.01)	Extra costs for amenities (0.17), extra charges (0.14), room with sea view (0.05), best stay ever (0.04), luxury expensive amenities (0.03), spa gym massage (0.03), amazing dinners (0.01), dirty room beddings and shower (0.0.01), bar food with cats (0.01)
Higher Power Distance	Smelly room (0.18), close to the beach (0.06), good location (0.04)	Extra charges (0.14), beach hotel with convenient location (0.06), spa gym massage (0.06), extra costs for amenities (0.04)
Lower Power Distance	Good entertainment (0.2), accept late arrivals (0.11), limited options of all inclusive (0.08), bad customer service (0.08), poor entertainment and food (0.05), ideal for summer holidays (0.05), good access e.g. bus stations (0.03), amazing staff (0.02), good bar service (0.02), comfortable room (0.01), basic apartment no luxuries (0.01), friendly staff (0.01)	Nice pool drinks (0.18), located close to the beach (0.17), amazing dinners (0.03), room with sea view (0.02), great team of staff (0.0.2), lovely staff (0.02), perfect location (0.01), missing services (0.01), bar food with cats (0.01)
Higher Uncertainty Avoidance	Confrontable room (0.21), accept late arrivals (0.17), good location (0.04), bad customer service (0.03), clean and good service (0.02), poor entertainment and food (0.01), amazing staff (0.01)	Extra charges (0.13), room with sea view (0.1), spa gym massage (0.09), luxury expensive amenities (0.05), disagree with hotels reviews (0.01)
Lower Uncertainty Avoidance	Ideal for summer holidays (0.12), good entertainment (0.07), low quality food and drinks (0.05), good access bus stations (0.04), close to beach (0.04), good bar service (0.02), limited options of all inclusive (0.02), basic accommodation (0.01)	Lovely staff (0.17), nice pool drinks (0.17), bar food with cats (0.03), good transportation options (0.02), located close to the beach (0.02), perfect location (0.02), extra costs for amenities (0.01)
Higher Individualism	Good entertainment (0.16), bad customer service (0.10), poor entertainment and food (0.08), limited option of all inclusive (0.01)	Nice pool drinks (0.18), great team of staff (0.16), pool area issues (0.08)
Lower Individualism	Accept late arrivals (0.19), good location (0.11), close to the beach (0.04), smelly room (0.03), comfortable room (0.02), will revisit (0.01)	Extra charges (0.14), spa gym massage (0.08), extra cost for amenities (0.06), luxury expensive amenities (0.06), located close to the beach (0.01), missing services (0.01), lovely staff (0.01), check in front desk issues (0.01), disagree with hotel reviews (0.01), beach hotel with convenient location (0.01), bar food cats (0.01), good transportation options (0.01)
Higher Masculinity	Good entertainment (0.17), good bar service (0.07), ideal for summer holidays (0.03)	Nice pool drinks (0.16), pool area issues (0.02)
Lower Masculinity	Accept late arrivals (0.18), close to the beach (0.1), smelly room (0.06), comfortable room 0.04(), old room (0.04), good location (0.02)	Extra charges (0.20), spa gym massages (0.09), luxury expensive amenities (0.07), extra cost for amenities (0.06), beach hotel with convenient location (0.06), lovely staff (0.02), room with sea view (0.01), bar food with cats (0.01)
Higher Long-Term Orientation	Good entertainment (0.18), ideal for summer holidays (0.06), low quality food and drinks (0.06)	Lovely staff (0.06), beach hotel with convenient location (0.01)
Lower Long-Term Orientation	Good location (0.15), close to the beach (0.12), accept late arrivals (0.06), comfortable room (0.05), smelly room (0.02), old room (0.01)	Spa gym massage (0.15), extra charges (0.15), luxury expensive amenities (0.05), best stay ever (0.05), room with sea view (0.03), extra costs for amenities (0.01), excellent service (0.01)
Higher Indulgence	Low quality food and drinks (0.19), poor entertainment and food (0.16), good entertainment (0.04)	Nice pool drinks (0.09), located close to the beach (0.03), missing services (0.02), dirty room beddings and shower (0.01)
Lower Indulgence	Accept late arrivals (0.11), good location (0.07), close to the beach (0.06), old room (0.02), smelly room (0.01), amazing staff (0.01)	Extra costs for amenities (0.17), extra charges (0.17), beach hotel with convenient location (0.08), luxury expensive amenities (0.05), spa gym massage (0.03), room with sea view (0.02), lovely staff (0.01)

capita), and satisfaction to support message design. The proposed method is applied to a Cyprus tourists' eWOM case study and the results are used to inform marketing messages' design for improved targeting.

The results show that guests from wealthier countries are relatively more demanding as hotel customers and make more complaints about extra costs when staying in 4/5-star hotels. The location criterion is affected the most in lower GDP per capita tourists. High power distance tourists who stay in 2/3-star hotels are likely to complain about "room smell" while higher power distance tourists that stay in 4/5 star hotels are likely to enjoy pool drinks. Higher uncertainty avoidance tourists that stay at 2/3-star hotels are likely to embrace room comfort while lower uncertainty avoidance that stay in 4/5-star hotels appreciate the good staff attitude and pool drinks. High individualism tourists staying at 4/5-star hotels are likely to embrace pool drinks while lower individualism tourist staying in 2/3-star hotels accept the hotel service during late arrivals. Higher masculinity tourists staying at 2/3-star hotels are likely to appreciate entertainment while lower masculinity tourists that stay in 4/5-star hotels to criticise extra charges. Higher long-term orientation tourists that stay in 2/3-star hotels are likely to embrace good entertainment while lower long-term orientation tourists that stay at 4/5-star hotels like the spa and gym facilities but complain for extra costs. Finally, higher indulgence tourists staying in 2/3-star hotels are likely to criticise food and drinks and low indulgence tourists that stay at 4/5-star hotels are likely to criticise extra costs for amenities and extra charges.

7. Management implications

The method can help hoteliers optimise their marketing campaigns by designing messages that could motivate target consumers by utilising the results in Table 1. Designing messages using topics from the identified patterns increases the likelihood of positively influencing the recipient of the message. Management can utilise identified patterns to target tourists through means such as, emails, social media, or dynamically adjusting the content in their website, to increase the probability of recipients engaging with the content and being motivated. The process initiates by evaluating the country's cultural and economic

Table 4
Comparative analysis of popular white box classification techniques against the DT for the two groups of hotels. In bold the model(s) with best AUC.

	2–3 Star Hotels AUC				4–5 Star Hotels AUC			
Target/output variable	**DT**	**BBN**	**Naïve Bayes**	**Ridge Classifier**	**DT**	**BBN**	**Naïve Bayes**	**Ridge Classifier**
Sentiment	**89**	**89**	**89**	81	**90**	**90**	88	86
GDP per capita	**78**	75	76	74	**79**	71	77	77
Power distance	**79**	77	77	77	77	72	**79**	77
Uncertainty avoidance	**79**	76	78	75	77	71	**78**	75
Individualism	75	75	**77**	73	**76**	65	74	75
Masculinity	76	76	**78**	76	**78**	75	76	76
Long-term orientation	**77**	75	**77**	74	**78**	75	**78**	76
Indulgence	**79**	78	78	78	79	69	**80**	77

conditions and subsequently uses these to locate topics based on their relation to the touristic destination. During the design of messages, the management should highlight aspects associated with positive topics and minimise or inverse the messages associated with negative topics. Management can also utilise the results highlighted from the topics-sentiment analysis to act on issues with regards to customers' dissatisfaction. Such actions can increase customers' intention to revisit, due to increased satisfaction, a property that requires less investment in contrast to new customers' attainment.

The novelty of this work resides in the combination of exogenous eWOM information derived from cultural gap and economic distance theory with endogenous eWOM's properties to inform tourist DCM's message design. Methodologically, the contribution resides in the use of interpretable models with good predictive performance, realised through the combination of multiple DTs optimised using different techniques. Results have shown that the proposed method outperform most of the other white box techniques that we examined.

8. Limitations and future directions

The main limitation of this work resides in the quality of the data and issues with possible fake reviews that might have affected the results. Future work can focus on filtering out these reviews using the recommendations of Liu and Pang (2018) to examine if these alter in any way the main conclusions of the current study. Another limitation concerns recent arguments about the selectivity of those reviewing; this is a timely issue at the moment and a current investigation topic within the larger community of big data analytics community.

Finally, we aim to evaluate further the methodology and toolset with more validation studies using comparative analysis of the decisions made by marketing managers against the suggestions of the method. This however constitutes a higher-level evaluation of the method at the technology readiness level scale and hence is intended at the next stage of our future work.

CRediT authorship contribution statement

Andreas Gregoriades: Conceptualization, Methodology, Formal analysis, Validation, Investigation, Writing - original draft, Supervision, Writing - review & editing. **Maria Pampaka:** Writing - original draft, Writing - review & editing, Formal analysis, Methodology. **Herodotos Herodotou:** Writing - original draft, Writing - review & editing, Formal analysis, Methodology. **Evripides Christodoulou:** Data curation, Software, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahn, M., & McKercher, B. (2015). The effect of cultural distance on tourism: A study of international visitors to Hong Kong. *Asia Pacific Journal of Tourism Research, 20*(1), 94–113. https://doi.org/10.1080/10941665.2013.866586

Alami, N., Meknassi, M., & En-nahnahi, N. (2019). Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. *Expert Systems with Applications, 123*, 195–211. https://doi.org/10.1016/j.eswa.2019.01.037

Aviera, G., Akinwale, A., & Fontecha, M. R. (2021). Machine-driven content marketing. In *The Machine Age of Customer Insight* (pp. 51–64). Emerald Publishing Limited: Bingley.

Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management, 53*, 125–131.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Blut, M., Teller, C., & Floh, A. (2018). Testing retail marketing-mix effects on patronage: A meta-analysis. *Journal of Retailing, 94*(2), 113–135. https://doi.org/10.1016/j.jretai.2018.03.001

Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual web texts. *Information Retrieval, 12*(5), 526–558. https://doi.org/10.1007/s10791-008-9070-z

Busacca, B., & Padula, G. (2005). Understanding the relationship between attribute performance and overall satisfaction. *Marketing Intelligence and Planning, 23*(6), 543–561.

Chang, Y.-C., Ku, C.-H., & Chen, C.-H. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. *Tourism Management, 80*, 104129. https://doi.org/10.1016/j.tourman.2020.104129

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Chopard, B., & Tomassini, M. (2018). Simulated annealing. In *Natural Computing Series* (pp. 59–79). https://doi.org/10.1007/978-3-319-93073-2_4

Christodoulou, E., Gregoriades, A., Pampaka, M., & Herodotou, H. (2020). Combination of topic modelling and decision tree classification for tourist destination marketing. In *Lecture Notes in Business Information Processing* (pp. 95–108). https://doi.org/10.1007/978-3-030-49165-9_9

Dipietro, W. R., & Anoruo, E. (2006). GDP per capita and its challengers as measures of happiness. *International Journal of Social Economics, 33*(10), 698–709. https://doi.org/10.1108/03068290610689732

Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM, 63*(1), 68–77. https://doi.org/10.1145/3359786

Dubois, B., & Duquesne, P. (1993). The market for luxury goods: Income versus culture. *European Journal of Marketing, 27*(1), 35–44. https://doi.org/10.1108/03090569310024530

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., … Wang, Y. (2020). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management, 59*, 102168. https://doi.org/10.1016/j.ijinfomgt.2020.102168

Epaminonda, E. (2020). A review of the sociocultural profile of Cyprus: Historical background, current features, change and diversity. *The Cyprus Review*.

Fischer, J., & Lipovska, H. (2018). Coffee index as quick and simple indicator of purchasing power parity. *Statistika, 98*(1), 55–67.

Frank, B., Enkawa, T., & Schvaneveldt, S. J. (2015). The role of individualism vs. collectivism in the formation of repurchase intent: A cross-industry comparison of the effects of cultural and personal values. *Journal of Economic Psychology, 51*, 261–278. https://doi.org/10.1016/j.joep.2015.08.008

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review, 47*(1), 1–66. https://doi.org/10.1007/s10462-016-9475-9

Gilboa, S., & Mitchell, V. (2020). The role of culture and purchasing power parity in shaping mall-shoppers' profiles. *Journal of Retailing and Consumer Services, 52*, 101951. https://doi.org/10.1016/j.jretconser.2019.101951

Gómez, S. E., Hernández-Callejo, L., Martínez, B. C., & Sánchez-Esguevillas, A. J. (2019). Exploratory study on class imbalance and solutions for network traffic classification. *Neurocomputing, 343*, 100–119. https://doi.org/10.1016/j.neucom.2018.07.091

Gregoriades, A., & Pampaka, M. (2020). Electronic word of mouth analysis for new product positioning evaluation. *Electronic Commerce Research and Applications, 42*, 100986. https://doi.org/10.1016/j.elerap.2020.100986

Gridach, M. (2020). Hybrid deep neural networks for recommender systems. *Neurocomputing, 413*, 23–30. https://doi.org/10.1016/j.neucom.2020.06.025

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Hofstede, G. (1980). *Cultural consequences: International differences in work-related values*. Beverly-Hills, CA: Sage Publication.

Hofstede, G., Hofstede, G., & Minkov, M. (2010). Cultures and organizations: Software of the mind. *Cultures and Organizations* (3rd ed.). McGraw Hill Professional.

Hollebeek, L. D., & Macky, K. (2019). Digital content marketing's role in fostering consumer engagement, trust, and value: Framework, fundamental propositions, and implications. *Journal of Interactive Marketing, 45*, 27–41. https://doi.org/10.1016/j.intmar.2018.07.003

Huang, J.-H., Huang, C.-T., & Wu, S. (1996). National character and response to unsatisfactory hotel service. *International Journal of Hospitality Management, 15*(3), 229–243.

Johnson, M. D., Herrmann, A., & Gustafsson, A. (2002). Comparing customer satisfaction across industries and countries. *Journal of Economic Psychology, 23*(6), 749–769. https://doi.org/10.1016/S0167-4870(02)00137-X

Jones, M. L. (2007). Hofstede - culturally questionable?

Kim, C. S., & Aggarwal, P. (2016). The customer is king: Culture-based unintended consequences of modern marketing. *Journal of Consumer Marketing, 33*(3), 193–201. https://doi.org/10.1108/JCM-01-2015-1273

Kim, J. M., Jun, M., & Kim, C. K. (2018). The effects of culture on consumers' consumption and generation of online reviews. *Journal of Interactive Marketing, 43*, 134–150. https://doi.org/10.1016/j.intmar.2018.05.002

Kirilenko, A. P., Stepchenkova, S. O., & Dai, X. (2021). Automated topic modeling of tourist reviews: Does the Anna Karenina principle apply? *Tourism Management, 83*, Article 104241. https://doi.org/10.1016/j.tourman.2020.104241

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of Culture's Consequences: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies, 37*(3), 285–320. https://doi.org/10.1057/palgrave.jibs.8400202

Koh, N. S., Hu, N., & Clemons, E. K. (2010). Do online reviews reflect a product's true perceived quality? - An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications, 9*(5), 374–385. https://doi.org/10.1016/j.elerap.2010.04.001

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(1–2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

Koller, D., & Friedman, N. (2009). Probabilistic graphical models: principles and techniques (Adaptive Computation and Machine Learning series). *Foundations*.

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications, 116*, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037

Kotler, P., & Armstrong, G. (2018). *Principles of marketing*. Pearson.

Laguna, M. (2018). Tabu search. In *Handbook of Heuristics* (pp. 741–758). https://doi.org/10.1007/978-3-319-07124-4_24

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature Selection. *ACM Computing Surveys, 50*(6), 1–45. https://doi.org/10.1145/3136625

Liu, J., Liao, X., Huang, W., & Liao, X. (2019). Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision. *Omega, 83*, 1–13. https://doi.org/10.1016/j.omega.2018.01.008

Liu, Y., & Pang, B.o. (2018). A unified framework for detecting author spamicity by modeling review deviation. *Expert Systems with Applications, 112*, 148–155. https://doi.org/10.1016/j.eswa.2018.06.028

López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250*, 113–141. https://doi.org/10.1016/j.ins.2013.07.007

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2016). An unexpected unity among methods for interpreting model predictions. *29th Conference on Neural Information Processing Systems. Barcelona, Spain*.

Luo, X., & Donthu, N. (2006). Marketing's credibility: A longitudinal investigation of marketing communication productivity and shareholder value. *Journal of Marketing, 70*(4), 70–91. https://doi.org/10.1509/jmkg.70.4.70

Madden, M. G. (2009). On the classification performance of TAN and general Bayesian networks. *Knowledge-Based Systems, 22*(7), 489–495. https://doi.org/10.1016/j.knosys.2008.10.006

Manosuthi, N., Lee, J.-S., & Han, H. (2020). Impact of distance on the arrivals, behaviours and attitudes of international tourists in Hong Kong: A longitudinal approach. *Tourism Management, 78*, 103–963. https://doi.org/10.1016/j.tourman.2019.103963

Mantovani, R. G., Horvath, T., Cerri, R., Vanschoren, J., & De Carvalho, A. C. P. L. F. (2017). Hyper-parameter tuning of a decision tree induction algorithm. In *Proceedings - 2016 5th Brazilian Conference on Intelligent Systems, BRACIS 2016*. https://doi.org/10.1109/BRACIS.2016.018

Matzler, K., Bailom, F., Hinterhuber, H. H., Renzl, B., & Pichler, J. (2004). The asymmetric relationship between attribute-level performance and overall customer satisfaction: A reconsideration of the importance–performance analysis. *Industrial Marketing Management, 33*(4), 271–277. https://doi.org/10.1016/S0019-8501(03)00055-5

Mazanec, J. A., Crotts, J. C., Gursoy, D., & Lu, L. (2015). Homogeneity versus heterogeneity of cultural values: An item-response theoretical approach applying Hofstede's cultural dimensions in a single nation. *Tourism Management, 48*, 299–304. https://doi.org/10.1016/j.tourman.2014.11.011

Morgeson, F. V., Mithas, S., Keiningham, T. L., & Aksoy, L. (2011). An investigation of the cross-national determinants of customer satisfaction. *Journal of the Academy of Marketing Science, 39*(2), 198–215. https://doi.org/10.1007/s11747-010-0232-3

O'Connor, P. (2008). User-generated content and travel: A case study on Tripadvisor. Com. In P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 47–58). Vienna: Springer Vienna. https://doi.org/10.1007/978-3-211-77280-5_5.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research, 50*, 67–83. https://doi.org/10.1016/j.annals.2014.10.007

Pulizzi, J. (2014). Epic Content Marketing.

Purnawirawan, N., Eisend, M., De Pelsmacker, P., & Dens, N. (2015). A meta-analytic investigation of the role of valence in online reviews. *Journal of Interactive Marketing, 31*, 17–27. https://doi.org/10.1016/j.intmar.2015.05.001

Radojevic, T., Stanisic, N., Stanic, N., & Davidson, R. (2018). The effects of traveling for business on customer satisfaction with hotel services. *Tourism Management, 67*, 326–341. https://doi.org/10.1016/j.tourman.2018.02.007

Razi, M., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications, 29*(1), 65–74. https://doi.org/10.1016/j.eswa.2005.01.006

Reisinger, Y. (2009). *International tourism: Cultures and behaviour. The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference*. Burington, MA: Elsevier.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning. *Expert Systems with Applications, 181*, 115111. https://doi.org/10.1016/j.eswa.2021.115111

Sann, R., Lai, P. C., & Liaw, S. Y. (2020). Online complaining behavior: Does cultural background and hotel class matter? *Journal of Hospitality and Tourism Management*. https://doi.org/10.1016/j.jhtm.2020.02.004

Schofield, A., & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics, 4*, 287–300. https://doi.org/10.1162/tacl_a_00099

Schuckert, M., Liu, X., & 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, 48*, 143–149. https://doi.org/10.1016/j.ijhm.2014.12.007

Shan, G., Xu, S., Yang, L.i., Jia, S., & Xiang, Y. (2020). Learn#: A Novel incremental learning method for text classification. *Expert Systems with Applications, 147*, 113198. https://doi.org/10.1016/j.eswa.2020.113198

Shapley, L. S. (1953). Stochastic Games. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.39.10.1095

Sharma, A., Woodward, R., & Grillini, S. (2020). Unconditional quantile regression analysis of UK inbound tourist expenditures. *Economic Letters, 186*, 108–857.

Sharpley, R. (2014). Host perceptions of tourism: A review of the research. *Tourism Management, 42*, 37–49. https://doi.org/10.1016/j.tourman.2013.10.007

Songshan, H., & John, C. (2019). Relationships between Hofstede's cultural dimensions and tourist satisfaction: A cross-country cross-sample examination. *Tourism Management, 72*, 232–241. https://doi.org/10.1016/j.tourman.2018.12.001

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics, 1*(1), 3–20. https://doi.org/10.1007/s11569-007-0005-8

Valdivia, A., Luzon, M. V., & Herrera, F. (2017). Sentiment analysis in tripadvisor. *IEEE Intelligent Systems, 32*(4), 72–77.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences, 22*(2), 219–240. https://doi.org/10.1111/j.1540-5915.1991.tb00344.x

Voosen, P. (2017). How AI detectives are cracking open the black box of deep learning. *Science*. https://doi.org/10.1126/science.aan7059

Wang, Y., Xiong, M., & Olya, H. (2020). Toward an Understanding of Responsible Artificial Intelligence Practices. In Proceedings of the 53rd Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2020.610.

Wang, Y. Y., & Li, J. (2008). Feature-selection ability of the decision-tree algorithm and the impact of feature-selection/extraction on decision-tree results based on hyperspectral data. *International Journal of Remote Sensing, 29*(10), 2993–3010. https://doi.org/10.1080/01431160701442070

Winter, S., Maslowsk, E., & Vosc, A. (2020). The effects of trait-based personalization in social media advertising. *Computers in Human Behavior, 114*, Article 106525.

Witten, I. H., Frank, E., & Hall, M. a. (2011). Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (3rd ed.). https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

Wong, J., & Law, R. (2003). Difference in shopping satisfaction levels: A study of tourists in Hong Kong. *Tourism Management, 24*(4), 401–410.

Yan, J., Zhang, Z., Lin, K., Yang, F., & Luo, X. (2020). A hybrid scheme-based one-vs-all decision trees for multi-class classification tasks. *Knowledge-Based Systems, 198*, 105922. https://doi.org/10.1016/j.knosys.2020.105922

Yen, C. L. A., & Tang, C. H. H. (2019). The effects of hotel attribute performance on electronic word-of-mouth (eWOM) behaviors. *International Journal of Hospitality Management, 76*, 9–18. https://doi.org/10.1016/j.ijhm.2018.03.006

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? *Information Technology & Tourism, 10*(4), 283–295.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications, 38*(6), 7674–7682. https://doi.org/10.1016/j.eswa.2010.12.147

Zhao, J., Jin, J., Chen, S., Zhang, R., Yu, B., & Liu, Q. (2020). A weighted hybrid ensemble method for classifying imbalanced data. *Knowledge-Based Systems, 203*, 106087. https://doi.org/10.1016/j.knosys.2020.106087

Zhou, H., Zhang, J., Zhou, Y., Guo, X., & Ma, Y. (2021). A feature selection algorithm of decision tree based on feature weight. *Expert Systems with Applications, 164*, 113842. https://doi.org/10.1016/j.eswa.2020.113842