Combination of Topic Modelling and Decision Tree Classification for Tourist Destination Marketing

Evripides Christodoulou[1], Andreas Gregoriades[1(✉)],
Maria Pampaka[2], and Herodotos Herodotou[1]

[1] Cyprus University of Technology, Limassol, Cyprus
ep.xristodoulou@edu.cut.ac.cy, {andreas.gregoriades,
herodotos.herodotou}@cut.ac.cy
[2] The University of Manchester, Manchester, UK
maria.pampaka@manchester.ac.uk

Abstract. This paper applies a smart tourism approach to tourist destination marketing campaigns through the analysis of tourists' reviews from TripAdvisor to identify significant patterns in the data. The proposed method combines topic modelling using Structured Topic Analysis with sentiment polarity, information on culture, and purchasing power of tourists for the development of a Decision Tree (DT) to predict tourists' experience. For data collection and analysis, several custom-made python scripts were used. Data underwent integration, cleansing, incomplete data processing, and imbalance data treatments prior to being analysed. The patterns that emerged from the DT are expressed in terms of rules that highlight variable combinations leading to negative or positive sentiment. The generated predictive model can be used by destination management to tailor marketing strategy by targeting tourists who are more likely to be satisfied at the destination according to their needs.

Keywords: Topic modelling · Sentiment analysis · Decision tree · Tourists' reviews

1 Introduction

With the recent information explosion as a result of the proliferation of data from social media, a new challenge emerged to discover information patterns hidden in big data using effective data mining techniques [19]. Micro-blogs are small messages communicated via social media such as Twitter, and gained popularity recently as means of expressing peoples' views [6]. Micro-blogs fall under the category of unstructured big data and are considered a type of electronic word of mouth (eWOM). A significant amount of eWOM are generated as part of consumers' evaluations of products and the hospitality services they are linked to [2, 33]. Hence, consumers and tourists now play an active role in shaping an organization's reputation [13], which in turn can impact the organisation's sales performance [36]. Therefore, the analysis of reviews has become a mainstream activity in marketing to improve product and services positioning based on

customers' needs and opinions [18]. According to [28], a brand is no longer what the company tells a customer it is, but rather, what customers tell each other it is.

TripAdvisor and other social media platforms have become valuable sources of eWOM in the tourism domain, with several studies investigating sentiment in reviews [27] given evidence that it can predict product success [29]. Topic modelling has also been used to identify topics discussed in reviews to provide temporal associations between topics in a timeline. These studies, however, concentrated on endogenous aspects of tourists' reviews (i.e., sentiment) whilst exogenous aspects, such as culture and purchasing power, have been addressed separately. No study so far combines exogenous with endogenous variables in one model to investigate the reasons for tourist dissatisfaction and predict perspective tourists' sentiment. Therefore, this paper investigates the application of Decision Trees to identify patterns, by evaluating the combined effect of culture, purchasing power, and topics discussed by tourists in reviews, on sentiment polarity. The research questions addressed in this paper are: (i) What are the main patterns emerging from tourist reviews of Cyprus hotels? (ii) How endogenous and exogenous reviews' parameters affect tourist sentiment?

The paper is organized as follows. Section 2 reviews the literature pertaining to the effect of culture and purchasing power on tourists' experience. Sections 3 and 4 elaborate on the proposed methodology and the obtained results, respectively. The paper concludes with the implications of the research and future directions.

2   Literature Review

The literature related to culture and purchasing power is presented next.

2.1   Culture

The driver to address culture within our research question is grounded on evidence that the tourists' cultural values, such as power distance, individualism, and uncertainty avoidance, significantly affect their perception of service quality, service evaluation, and satisfaction [20] Other studies indicate that the customers' power distance significantly affects their service expectations, perceived service quality, and relationship quality [10]. A key factor that differentiates tourist activities at a destination is culture, with studies, e.g. [8], identifying that certain traits have significant differences. This theory is also supported in consumer behaviour by evidence showing that people of the same nationality tend to have similar preferences [16].

There are several models of culture. In this study, we adopted the model of Hofstede [14] due to its eminent reputation. According to this model, there are six different traits that form a culture: (1) *Power Distance* (i.e. the degree to which people accept and expect that power is distributed unequally); (2) *Individualism* (i.e. the degree to which people tend to take care of only themselves and their immediate families); (3) *Masculinity* (i.e. the degree to which achievement, heroism, assertiveness, and material rewards for success are preferred); (4)*Uncertainty Avoidance* (i.e. the degree to which risk and uncertainty tend to be avoided; (5) *Long Term Orientation* (i.e. the degree to which people prefer stability, respect for tradition, and are future-oriented);

and (6) *Indulgence* (i.e. the degree to which people prefer freedom and free will). For the purpose of this study, we used Hofstede's cross-cultural differences model (similar to [16]) to obtain each reviewer's culture values to enhance our tourist review data.

2.2 Purchasing Power

The use of purchasing power is grounded on evidence highlighting that customers from countries with greater power distance feel superior to service providers [20] and expect high service quality. This is linked to evidence that purchasing power [38] is linked with a greater need to portray status through consumption [12], hence promoting power distance. The financial state of a country has been used for global markets analysis [15] with Gross Domestic Product (GDP) per capita as a key indicator for comparing the level of development among countries and socioeconomic status. Human welfare and GDP per capita go together, while increased GDP per capita is found to be correlated with happiness [11]. At the same time, in countries with low human development index, GDP dramatically affects quality of life [17]. Therefore, the argument by many researchers is that tourists from countries with lower purchasing power compared to their tourist destination might be more demanding and hence more likely to evaluate their experience at a destination negatively.

3 Methodology

The techniques used to address our research questions include sentiment analysis, topic modelling, decision trees, and imbalance data treatment. These are described in turn along with the overall proposed methodology that combines them.

3.1 Sentiment Analysis

Sentiment analysis (SA) and opinion mining have been studied and used for a while with several techniques emerging for analysing emotions and opinions from eWOM [26]. SA is useful for online opinions analysis due to its ability to automatically measure emotion in online content using algorithms to detect polarity in eWOM [32]. Three common SA approaches are Machine Learning (ML), Lexicon-based Methods, and Linguistic Analysis techniques. From these ML techniques are considered the most effective and simplest to use with Naïve Bayes (NB) and Support Vector Machines (SVM) being the most popular. ML techniques are classified into supervised and unsupervised [41], with supervised requiring training the classifier prior to its use. The main difference from unsupervised is that supervised techniques use labelled opinions that have been pre-evaluated as negative, positive, or neutral to train models. Such techniques include SVM, NB, Logistic Regression, Multilayer Perceptron, K-Nearest Neighbours, and Decision Trees [23]. In this study, the NB approach is employed for SA due to its good results and popularity [3, 24, 39].

3.2   Topic Modelling

Topic modelling, a type of unsupervised data mining technique, constitutes a popular tool for extracting important themes (topics) from unstructured data and is employed to reveal and annotate large documents collection with thematic information [31]. Two of the most popular techniques for topic analysis are the Latent Dirichlet Allocation (LDA) and the Structural Topic Model (STM) [14]. In LDA, a topic is a probability distribution function over a set of words used as a type of text summarization. LDA expresses the relationships between words in terms of their affinity to certain latent variables (topics), using Bayesian probabilities. STMs extend the LDA framework with the capability of accommodating supplementary information in the form of metadata that could reveal important aspects of how topics are linked to covariates [35], or to observe which topics correlate with one another [9]. LDA and STM are generative models and assume that each topic is a distribution over words and each document is a mixture of topics [7]. STM is employed in this study, with each review representing a distribution of a finite set of topics, which in turn are distributions of the words in the corpus used in similar reviews. Identified topics were later associated to each review in the dataset. New columns are added depending on the number of topics, each representing the degree of association of each topic to the case.

3.3   Decision Trees

Decision Trees (DT) are considered a scalable multivariate method and have been successfully applied in prediction problems by mimicking the human decision-making process. They are intuitive and explanatory, unlike black-box algorithms such as support vector machines or artificial neural networks that cannot be easily comprehended by decision makers or validated by domain experts. A DT learns its structure by partitioning the training dataset into bins using a series of splits, each performed after identifying the most prevalent split-variable using information gain or Gini impurity index metrics. Splitting variables are used in defining the structure of the tree that is made up of nodes. Each node splits the dataset into branches. There are several algorithms for designing DTs such as CART, ID3, C4.5, CHAID etc. The CART algorithm [5] is a binary classification technique that utilises the Gini index of heterogeneity to determine the information gain of each variable and accordingly decide which variables to be used to split the dataset.

The main advantages of DTs lie in their simple interpretation and visualization capabilities, and the need for little data preparation. They can handle both numerical and categorical data. Their drawbacks include creation of over-complex trees that could sometimes overfit the data and not generalize well. This is due to the use of heuristic algorithms such as the greedy algorithm, where locally optimal decisions are made at each node. Another drawback is that DT algorithms create biased trees if the training set is imbalanced (large difference in number of cases representing the class variable). It is, thus, recommended to balance the dataset as explained next, prior to DT training.

Hyperparameter tuning is another activity performed prior to model learning to find the optimum configuration of a DT for improved model performance. There is no uniform way to specify hyperparameter values to reduce the loss of model

performance; experimentation through a grid-based search is a common approach. However, supervised algorithms can be used to automate this process. The main parameters utilized in this study to optimize the performance of the DT was the alpha value (DT cost complexity), the DT maximum depth, and minimum samples per leaf node [25].

3.4 Data Balancing

Data imbalance refers to the situation when the minority class of a dataset is much smaller than the majority class. In our case, the number of positive sentiment reviews was much larger than the negative ones (minority). This class imbalance can mislead the classifier into overfitting, since the majority class dominates the dataset. Hence, the classifier always generates results that abide with the majority class. Solutions to the class-imbalance problem include many different forms of under-sampling or over-sampling. The oversampling approach creates a balanced subset from the original dataset by duplicating samples of the minority class. Two of the most common oversampling techniques are Random oversampling (RO) and the Synthetic Minority Oversampling Technique (SMOTE). RO is easy to implement and involves the minority samples in the data being replicated randomly until the proportion of majority class is achieved. The SMOTE technique generates artificial samples from the minority class by combining several minority-class instances that are similar. That is, for each minority instance, it introduces a synthetic new sample by utilizing information from the minority-class nearest-neighbours instances. SMOTE is a more sophisticated technique and generally produces better results [40] than RO. We implemented and tested both approaches, confirming SMOTE's superior performance. Hence, SMOTE was employed to balance the training dataset before generating the DT model.

3.5 Overall Methodology

The main steps required to answer our research questions are depicted in Fig. 1. The first step is the collection of reviews, in English, from tourists who visited hotels in Cyprus during the period 2009-2019. This period is selected due to availability of data. The total number of reviews obtained from the data collection is 65000 from tourists representing 27 countries with the majority (85%) of cases representing years 2014-2019 due to the recent increase in eWOM popularity. Data was automatically extracted from TripAdvisor, with an algorithm developed in python that scrapped the reviews, and included: Username, Rating of hotel, Date of stay, Feedback date, Country of origin, Past contributions, Confidence votes, Review. To estimate each country's purchasing power, the GDP per capita index was used, using data from the World Monetary Fund. The variable is expressed in US dollars and was standardized in a scale from 0 to 100. Similarly, for the cultural values of each reviewer, the Hofstede website was used, associating each cultural dimension to a value in a scale from 0-100, based on country of origin. The tourists' reviews, GDP, and culture data were integrated to form a collated dataset used for DT training.

Prior to DT training, the data underwent cleansing, dimensionality reduction, and irrelevant data elimination. Cases with missing values (i.e., culture values) were
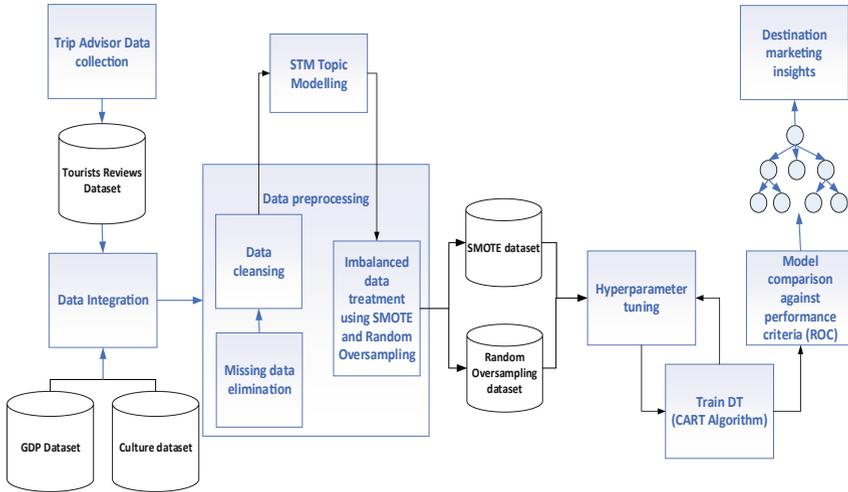
Fig. 1. Overall methodology

eliminated from the dataset, reducing the number of cases to 45000. The next step involved the analysis of consumers' sentiment and the topics discussed in the reviews, through polarity detection and topic analysis, respectively.

For the sentiment analysis, a pre-trained NB classifier was used, to evaluate the polarity of reviews initially in three categories: positive, negative, and neutral. The rationale for using a sentiment classifier instead of the actual review's rating in Trip Advisor, lies in evidence [22] suggesting that reviewers tend to refrain from giving low scores to hotels unless their experience is extremely negative. In this study [22] for instance they examined whether reviewers from collectivist-leaning societies (valuing tradition and helping each other) tend to write fewer excessively negative reviews than those from individualistic societies. They found that despite being marginally positive in ratings, reviewers use negative connotations in narrative descriptions that indicate dissatisfaction. To avoid this problem, we opted for a sentiment classification approach rather than using the reviewer's ratings alone. The two trained sentiment models used for this task were Textblob (based on Naive Bayes) and Vader [17], which are popular classifiers with satisfactory precision and recall scores. Both models were used in an ensemble manner to improve our confidence in the results. A python script automatically utilized the Textblob and Vader models and averaged their results. The process was repeated for all downloaded reviews, and their polarity was saved next to each review as a new attribute.

To reduce the data imbalance (5% negative, 10% neutral, and 85% positive), the sentiment polarity was converted to a binary state by merging the neutral and negative sentiments. This assumes that neutral sentiment is closer to the negative than the positive class due to reluctance of people giving negative reviews. This resulted in 15-85 data distribution, which was still imbalanced and was tackled in a subsequent step.

An additional issue that needed addressing prior to DT model learning was the fact that many of the reviews were from UK tourists. These represented 70% of the dataset

that monopolised the GDP and culture scores. Therefore, to minimize the bias from this majority group, a random sample was selected from the UK group to equate the maximum number of reviews from other tourists' groups. The resulting dataset included a total of 10,3K reviews.

The STM topic modelling approach is subsequently used to associate reviews with key thematic topics identified from the whole dataset. To learn the topic model, reviews had to be pre-processed further. During this step, irrelevant information was eliminated through the sub-steps of stop-word removal, tokenization, stemming. Stop-words refer to words providing little or no useful information to text analysis and can hence be considered as noise. Common stop-words include articles, conjunctions, prepositions and pronouns. Tokenization refers to the transformation of a stream of strings into a stream of processing units, referred to as tokens. Thus, during this step reviews were converted into a sequence of tokens, by choosing n-grams (phrases composed by n words in length). The Stemming process involves converting words to their root form. After data pre-processing, the STM topic modelling approach was employed. Extracted topics were inspected based on prior domain knowledge; therefore, expertise in the field under investigation was required to make the necessary connections. The identification of the recommended number of topics (k) was based on the model's semantic coherence and exclusivity for each model and topic [35]. The recommended value for k from this process was 10. Manual inspection of the resulting topics followed to identify possible topic merges into super-topics to reduce the dimensionality of the model further. This yielded the six super-topics of Table 2. Each review was then associated with super-topics based on the results of the trained model and the topic merges. The super-topic associations were appended as new variables in the datafile based on the probability distribution of topics per review, i.e., each review is associated with more than one topic.

To tackle the data imbalance challenge, the dataset that emerged from the super-topic assignment underwent treatment using SMOTE and RO techniques, yielding two new datasets that were utilized during DT training. Hyperparameter tuning was performed to identify the best configuration of the DT learning algorithm (CART) to maximize the model's performance. To validate the generated DTs from the SMOTE and RO datasets, the K-fold cross-validation approach was used. The Area Under the Receiver Operating Characteristic (ROC) curve (AUC), used for binary classification problems such as this one (pos/neg sentiment), describes the performance of a model as a whole and is useful for evaluating models trained on imbalanced data [4]. The higher the area under the ROC curve the better the model's performance. The best model was obtained using the SMOTE dataset with AUC 81.9%, while the DT generated from the RO dataset was inferior (78%) due to replicating existing cases from minority class.

The learned DT is finally used to identify the most significant patterns in the dataset. These are expressed as rules that combine both exogenous and endogenous variables of tourists' reviews and are indicated as nodes on the tree. Nodes also provide information concerning the polarity and are color-coded accordingly. The key rules are subsequently used to filter cases from the dataset that satisfy the rules, which are then used to estimate the distribution of tourists by country of origin. Countries with higher probabilities in each rules' sample-set, represent tourist origins that are more likely to be satisfied by existing services at a tourist destination. Such information can provide

interesting marketing insights to be used by destination managers to identify the things they do well or badly, and accordingly tailor their campaigns knowing which groups they can satisfy better.

4   Results

The first analytical step was the identification of the main topics discussed in the corpus of tourists' reviews. The STM method was used, and the recommended number of topics was identified to be ten based on the topic coherence, which denotes whether words in the same topic make sense when they are put together. The 10 topics identified are depicted in Table 1 along with the distribution of words for each topic. The table shows the most popular words that comprise each topic using different metrics such as highest probability, FREX, Lift, and Score. FREX weights words by their overall frequency and how exclusive they are to the topic, while Lift weights words by giving higher weight to words that appear less frequently in other topics. Score divides the log frequency of the word in the topic by the log frequency of the word in other topics. Based on these scores, words are presented in order from left to right, indicating their importance to the topic. The right column presents the interpretation of the topic in domain specific language.

To minimize the complexity of the topic model, the generated topics were merged into six super-topics based on common themes specified in Table 2. The last column in Table 2 provides the meaning of a high score for a super-topic. For example, a high cleanliness score indicates many complains about the cleanness of the hotel, while a high services/staff score indicates high satisfaction with the staff.

The DT classification task using the compiled data revealed that the combined use of the three predictor variables (culture, GDP per capita, super-topics) were able to predict group membership (reviews' sentiment polarity) better than chance (50/50) with overall prediction accuracy of 80%. The optimal decision tree is presented in Fig. 2. Each node depicts the state of each of the variables with super-topics expressed as a percentage with zero indicating no discussions about a topic and 100 many discussions. The values in each tree node denote the number of cases from the dataset that satisfy a given node criterion. The values on the left correspond to negative sentiment cases and the values on the right with positive. The Gini index (calculated by subtracting the sum of the squared probabilities of each class from one) shows the level of impurity of the distribution for each node with lower index indicating a larger difference between negative and positive sentiment. This is also visually represented using colour coding for the tree nodes. The Gini index is utilised to identify strong patterns in the dataset that are expressed in the form of rules.

The dominating variable in this tree is the topic associated with cleanliness followed by service and staff professionalism. High cleanliness issues (cleanliness > 20.5), indicating a lot of complains about cleanliness, yield 91% negative reviews (computed based on negative cases over total cases on the leaf node). This abides with results from [37]. However, when mild cleanliness is combined with high professional staff attitude (Service/Staff > 12.5), the reviews are slightly positive

Table 1. Emerging topics from STA and their interpretation

Topic	Top words per topic from STA based on four metrics	Interpretation
1	Highest Prob: room, hotel, bathroom, one, bed, shower, water FREX: shower, toilet, smell, window, eggs, air, bathroom Lift: toast, electric, scrambled, curtain, soap, ham, conditioner Score: toast, room, shower, bathroom, air, toilet, dirty	Cleanness of hotel and rooms
2	Highest Prob: hotel, staff, great, service, will, time, best FREX: wonderful, thank, amazing, thanks, vacation, highly, come Lift: die, wonderfull, pampered, costas, thanks, thank, surely Score: die, amazing, hotel, thank, great, staff, perfect	Service/Staff
3	Highest Prob: kids, restaurant, pool, well, pools, every, day FREX: resort, kids, club, gardens, chef, activities, adults Lift: elias, tranquility, exotic, azia, treatments, slides, dome Score: elias, kids, resort, team, pools, children, spa	Family friendly
4	Highest Prob: good, breakfast, hotel, room, rooms, clean, staff FREX: breakfast, comfortable, modern, view, spacious, rooms, wifi Lift: wonderfully, modern, satisfactory, spread, amphora, comfortable Score: wonderfully, good, breakfast, room, rooms, clean, comfortable	Breakfast and overall quality
5	Highest Prob: place, stay, stayed, cyprus, apartment, night, everything FREX: apartment, owner, apartments, place, accommodation, budget Lift: scale, asty, pyramos, san, remo, owner, apt Score: scale, place, apartment, apartments, stay, nicosia, owner	Budget stay
6	Highest Prob: hotel, food, dinner, get, day, evening, drinks FREX: board, half, drinks, eat, bay, lunch, inclusive Lift: famagusta, casino, brits, salamis, ajax, potatoes, Turkish Score: famagusta, half, entertainment, euro, board, food, inclusive	All-inclusive food
7	Highest Prob: hotel, room, day, didnt, one, got, night FREX: told, said, wasnt, didnt, asked, got, never Lift: paint, veronica, upset, sorted, complaining, dirt, realised Score: paint, told, didnt, said, asked, never, got	Complains about room

Table 1. (*continued*)

Topic	Top words per topic from STA based on four metrics	Interpretation
8	Highest Prob: hotel, walk, bus, minutes, beach, restaurants, larnaca FREX: walking, distance, bus, street, shops, walk, minutes Lift: route, mackenzie, cafes, intercity, lazarus, sights, mall Score: route, bus, walk, street, distance, city, larnaca	Hotel location and things to do
9	Highest Prob: hotel, guests, hotels, star, one, well, area FREX: guests, seasons, standards, four, management, star, years Lift: miserable, allows, result, wonâ, greatly, thoughtful, post Score: miserable, guests, seasons, star, years, four, management	Other guests
10	Highest Prob: nice, pool, beach, hotel, good, really, great FREX: napa, nice, ayia, pool, really, swimming, plaza Lift: pleasent, swiming, plaza, nestor, napa, ayia, aya Score: pleasent, nice, pool, napa, beach, great, good	Beach hotel and pool

Table 2. Super-topics that emerged from topic combinations and score interpretations

Topics with overlapping themes	Interpretation	High score meaning
Topic 1,7	Cleanliness	Complains
Topic 3,10, 6, 4	Family/Pool/Food	Satisfaction
Topic 5	Budget	Budget hotel
Topic 8	Location	Good location
Topic 2	Service/Staff	Satisfaction
Topic 9	Other Guests	High mixture of guest cultures

(67%). Staff professionalism and quality of service is a strong predictor of positive reviews with a weight of 93%, result also supported by [37].

Issues with multicultural guests (other guests) emerge with hotel guests from different cultures interacting in a hotel, combined with low staff satisfaction, leading to 81% negative reviews. An important cultural dimension that yields negative sentiment when the hotel service is not adequate is Power distance with a 65% weight on negative reviews. Low indulgence (indicating control of desires) yields positive sentiment in 71% of the cases when no cleanliness and staff issues are encountered, while the combination of high indulgence with low GDP per capita results in marginally negative sentiment, indicating that tourist from poorer countries who want to satisfy their desires
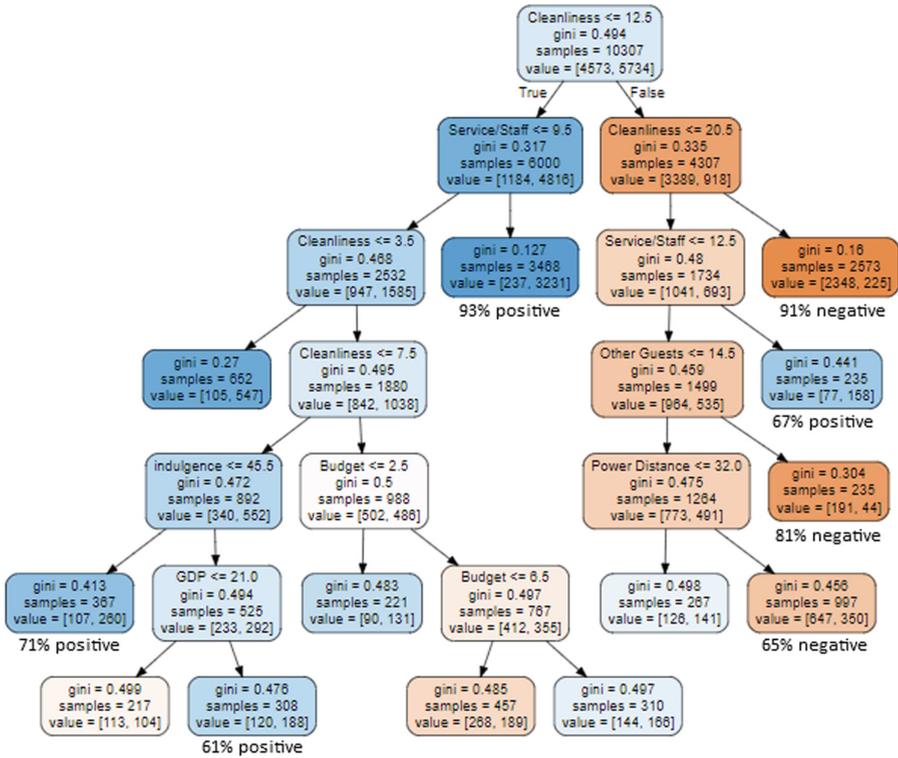
Fig. 2. Learned Decision Tree with colour coded sentiment predictions (negative shown in shades of orange and positive in shades of blue). Key results are annotated under the leafs. (Color figure online)

expect more for their money. On the other hand, tourists from richer countries tend to give positive sentiment in 61% of cases. Preliminary evaluation of the model's most prevalent patterns was performed using comparative analysis with relevant literature. Studies on cleanliness, service [37], and value for money (GDP) [30] report a similar effect on satisfaction. The latter is similar to our GDP dimension and utilises equity theory [1] to observe that satisfaction is occurring when the performance of the service equates the money paid for the service. The model was further evaluated by two experts from the hospitality industry who verified the validity of these rules.

5   Conclusions

This study serves as a proof-of-concept and is the first to combine a DT approach with topic modelling to identify patterns using both exogenous and endogenous parameters of tourists' reviews. Several studies examined mainly the impact of culture on review generation [21, 22, 34], therefore this work provides a contribution by combining other exogenous and endogenous variables. The case study used to elaborate this method

refers to Cyprus as a popular tourist destination and the data collected span the years 2009-2019. The results indicate that the three determinants of sentiment polarity in reviews are firstly, issues with cleanliness and staff professionalism that emerge from topic analysis, and then the cultural dimensions of indulgence and power distance. Based on the most prevalent rules from the DT and the cases that fall under each one, Cypriot hotels seem to have failed to satisfy tourists from Romania and Greece when issues with cleanliness emerge and the Power distance is relatively high. In contrast, when issues with Cleanliness alone emerge, tourists from the UK are more likely to be dissatisfied. Current hotel services cannot satisfy adequately tourists from these countries and, hence, destination marketing managers need to either improve their service or concentrate on tourists from other countries such as Australia, Switzerland and Netherlands that have higher GDP than Cyprus or have low levels of indulgence (<45) and are more likely to be satisfied. Tourist origin countries with positive results with regards to service quality and staff are, Israel, Lebanon and Greece, meaning that staff attitude and professionalism can affect their overall sentiment by 12% and 8% (Israel, Greece/Lebanon) compared to tourists from other countries with much lower positive sentiments. The main findings align with evidence from other studies indicating that consumers from countries with lower purchasing power provide low ratings to hotels. This is also consistent with evidence that power distance affects reviews polarity, supported by theory highlighting that in countries with high power distance, consumers often feel superior to service providers in the social hierarchy [20] and less tolerant with service quality, while they tend to give low service evaluations. Results from this work also highlight that other cultural traits from Hofstede, such as individualism, tend to be related to tourist review sentiment, while the topics that are associated with highest sentiment are those about hotel services. Limitations of this work reside in the quality of the data collected and issues pertaining to fake reviews that might affect the results. Our future work aims to filter out these reviews and examine if the effect of the aforementioned variables alters in any way the main conclusions of the current study.

References

1. Adams, J.S.: Towards an understanding of inequity. J. Abnorm. Soc. Psychol. **67**(5), 422 (1963)
2. Berger, J., Milkman, K.: What Makes Online Content Viral? J. Mark. Res. **49**, 192–205 (2012)
3. Boiy, E., Moens, M.F.: A machine learning approach to sentiment analysis in multilingual web texts. Inf. Retr. **12**(5), 526–558 (2009)
4. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. Patt. Recogn. **30**(7), 1145–1159 (1997)
5. Breiman, L., Friedman, J., Stone, C., Olshen, R.: Classification and Regression Trees (Wadsworth Statistics/Probability). CRC Press, New York (1984)
6. Chamlertwat, W., Bhattarakosol, P., Rungkasiri, T., Haruechaiyasak, C.: Discovering consumer insight from Twitter via sentiment analysis. J. Uni. Comput. Sci. **18**, 973–992 (2012)

7. Chaney, A.J.B., Blei, D.M.: Visualizing topic models. In: ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (2012)
8. Crotts, J., Erdmann, R.: Does national culture influence consumers' evaluation of travel services? A test of Hofstede's model of cross-cultural differences. MSQ **10**, 410–419 (2000)
9. Csardi, G., Nepusz, T.: The igraph software package for complex network research. Inter J. Complex Syst. **1695**, 1–9 (2006)
10. Dash, S., Bruning, E., Acharya, M.: The effect of power distance and individualism on service quality expectations in banking: a two-country individual- and national-cultural comparison. Int. J. Bank Mark. **27**, 336–358 (2009)
11. Dipietro, W.R., Anoruo, E.: GDP per capita and its challengers as measures of happiness. Int. J. Soc. Econ. **33**, 698–709 (2006)
12. Dubois, B., Duquesne, P.: The market for luxury goods: income versus culture. Eur. J. Mark. **27**, 35–44 (1993)
13. Etter, M., Ravasi, D., Colleoni, E.: Social media and the formation of organizational reputation. Acad. Manag. Rev. **44**, 28–52 (2019)
14. Gambhir, M., Gupta, V.: Recent automatic text summarization techniques: a survey. Artif. Intell. Rev. **47**, 1–66 (2017)
15. Gilboa, S., Mitchell, V.: The role of culture and purchasing power parity in shaping mall-shoppers' profiles. J. Retail. Consum. Serv. **52** (2020)
16. Huang, S., Crotts, J.: Relationships between Hofstede's cultural dimensions and tourist satisfaction: a cross-country cross-sample examination. Tour. Manag. **52**, 232–241 (2019)
17. Hutto, C.J., Gilbert, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: ICWSM (2014)
18. Jung, J.J.: Taxonomy alignment for interoperability between heterogeneous virtual organizations. Exp. Syst. Appl. **34**, 2721–2731 (2008)
19. Khade, A.A.: Performing customer behavior analysis using big data analytics. Procedia Comput. Sci. **79**, 986–992 (2016)
20. Kim, C.S., Aggarwal, P.: The customer is king: culture-based unintended consequences of modern marketing. J. Consum. Mark. **33**, 193–201 (2016)
21. Kim, J.M., Jun, M., Kim, C.K.: The effects of culture on consumers' consumption and generation of online reviews. J. Interact. Mark. **43**, 134–150 (2018)
22. Koh, N.S., Hu, N., Clemons, E.K.: Do online reviews reflect a product's true perceived quality? an investigation of online movie reviews across cultures. ECRA **9**, 374–385 (2010)
23. Krouska, A., Troussas, C., Virvou, M.: Comparative evaluation of algorithms for sentiment analysis over social networking services. J. Univ. Comput. Sci. **23**, 755–768 (2017)
24. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. (2012)
25. Mantovani, R.G., Horvath, T., Cerri. R., Vanschoren, J., De Carvalho, A.C.P.L.F.: Hyper-parameter tuning of a decision tree induction algorithm. In: BRACIS (2016)
26. Martin-Domingo, L., Martín, J.C., Mandsberg, G.: Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). J. Air. Transp. Manag. **78**, 106–115 (2019)
27. Moon, S., Kamakura, W.A.: A picture is worth a thousand words: translating product reviews into a product positioning map. Int. J. Res. Mark. **34**, 265–285 (2017)
28. Nayab, G., Bilal, M., Shrafat, A.: A brand is no longer what we tell the customer it is - it is what customers tell each other it is: (Lahome). Sci. Int. **28**, 2725–2729 (2016)
29. Nguyen, H., Chaudhuri, M.: Making new products go viral and succd. IJRM **36**, 39–62 (2019)
30. Nicolau, J.L., Mellinas, J.P., Martín-Fuentes, E.: Satisfaction measures with monetary and non-monetary components: hotel's overall scores. Int. J. Hosp. Manag. **87**, 102497 (2020)
31. Nikolenko, S.I., Koltcov, S., Koltsova, O.: Topic modelling for qualitative studies. J. Inf. Sci. **43**, 88–102 (2017)

32. Pang, B., Lee, L.: Opinion mining and sentiment analysis. FT Inf. Retr. **2**, 1–135 (2008)

33. Pfeffer, J., Zorbach, T., Carley, K,M.: Understanding online firestorms: negative word-of-mouth dynamics in social media networks. J. Mark. Commun. **20**, 117–128 (2014)

34. Purnawirawan, N., Eisend, M., De Pelsmacker, P., Dens, N.: A meta-analytic investigation of the role of valence in online reviews. J. Interact. Mark. **31**, 17–27 (2015)

35. Roberts, M., et al.: Structural topic models for open-ended survey responses. AJPS **58**, 1064–1082 (2014)

36. Rosario, A., Sotgiu, F., De Valck, K., Bijmolt, T.: The effect of eWOM on sales: a meta-analytic review of platform, product, and metric factors. J. Mark. Res. **53**, 297–318 (2016)

37. Sann, R., Lai, P.C., Liaw, S.Y.: Online complaining behavior: does cultural background and hotel class matter? J. Hosp. Tour. Manag. **43**, 80–90 (2020)

38. Schaninger, C.M.: Social class versus income revisited: an empirical investigation. J. Mark. Res. **18**, 192–208 (1981)

39. Singh, V.K., Piryani, R., Uddin, A., Waila, P.M.: Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches (2013)

40. Thabtah, F., Hammoud, S., Kamalov, F., Gonsalves, A.: Data imbalance in classification: experimental evaluation. Inf. Sci. (Ny) (2020)

41. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, Amsterdam (2011)