

RESEARCH

Open Access



Explaining tourist revisit intention using natural language processing and classification techniques

Andreas Gregoriades^{1*}, Maria Pampaka², Herodotos Herodotou¹ and Evripides Christodoulou¹

*Correspondence:
andreas.gregoriades@cut.ac.cy

¹ Cyprus University of Technology, Limassol, Cyprus
² The University of Manchester, Manchester, UK

Abstract

Revisit intention is a key indicator of business performance, studied in many fields including hospitality. This work employs big data analytics to investigate revisit intention patterns from tourists' electronic word of mouth (eWOM) using text classification, negation detection, and topic modelling. The method is applied on publicly available hotel reviews that are labelled automatically based on consumers' intention to revisit a hotel or not. Topics discussed in revisit-annotated reviews are automatically extracted and used as features during the training of two Extreme Gradient Boosting models (XGBoost), one for each of two hotel categories (2/3 and 4/5 stars). The emerging patterns from the trained XGBoost models are identified using an explainable machine learning technique, namely SHAP (SHapley Additive exPlanations). Results show how topics discussed by tourists in reviews relate with revisit/non revisit intention. The proposed method can help hoteliers make more informed decisions on how to improve their services and thus increase customer revisit occurrences.

Keywords: Revisit intention, Text classification, Negation detection, Topic modelling, Explainable machine learning

Introduction

Accommodation is a vital part of tourists' experience at a holiday destination and is associated with a complex decision-making process affected by tourists' motivations, service quality and venues' physical characteristics [1]. This process, however, is different for first time and repeating visitors with the latter offering many benefits to businesses as they converge into a decision faster and with less consumer acquisition costs [2], have more realistic expectations, stay longer, are satisfied more easily, tolerate service errors, and are more likely to spread positive electronic word of mouth (eWOM) [3]. Similarly, repurchase intentions expressed in eWOM are very persuasive and influence significantly other consumers' attitudes towards the service or product, which in turn increase sales [4]. Given these positive effects of revisits to businesses, stakeholders are keen to understand the causes of this consumer behaviour and accordingly improve their services to address consumers' needs [5]. Revisit intention and its drivers have become more important during the covid pandemic, due to the hotels' existential risk

caused by the drastic reductions in people's mobility [6]. The pandemic also highlighted the need for timely identification of these factors from eWOM through the application of machine learning (ML) and natural language processing (NLP) techniques, which in contrast to traditional methods, like questionnaires/surveys, enable businesses to swiftly react to new trends and reduce the risks associated with late interventions [7].

Revisit intention has been studied at the level of countries, cities, hotels, or attractions [3]. However, despite the importance of repeat visitors, there is no consistent and automated way to measure revisit intention, with the majority of studies employing questionnaires with different number of questions and drawing on various theoretical models. The most commonly used model to explore potential causes of repurchase involve the Theory of Planned Behavior (TPB) [8, 9], stimulus organism response theory [10], the value-attitude-behavior theory [11], and the consumer psychology theory [12]. However, the dependence of such methods on questionnaires makes them expensive, time consuming, and prone to social desirability bias, sampling bias, or other response biases (e.g., misunderstanding of questions by participants). To remain competitive, organisations require a timely analysis of consumers' behaviours and perceptions as well as automation support for the generation of actionable recommendations that can increase revisits and positive eWOM.

EWOM is the most popular way for consumers to seek and share information about products and services [13] and provides businesses with an alternative method to surveys for analysing consumers' behaviour in an automated way. The advantage of eWOM is that it enables customers to provide information about their experience without pressure or constraints from research instruments, and thus could provide an unbiased source of information and enable deciphering causes of revisit/non revisit intention [14–16] and shape consumers purchasing decisions [13]. In addition, eWOM enables the extraction of tacit knowledge that is considered superior to customer surveys, and is perceived by consumers as more credible than the content of official websites of hotels or destinations [17]. Many touristic organizations are, therefore, switching from collecting consumer views through questionnaires to mining eWOM to gauge customer satisfaction [18, 19], customer loyalty, and re-patronage [20]. Special attention is given by businesses to the customers/consumers who explicitly state their intention not to revisit in eWOM (i.e., non-revisitors). The timely response to such negative intentions is critical since they exert a greater impact on consumers decision-making than eWOM expressing positive revisit intentions [21].

Despite the need for timely analysis of revisiting, there is still limited research with applications of ML and NLP in revisit eWOM analysis and on how to leverage revisit causes to maximise profits [22]. Most previous studies focus on identifying revisit intention in eWOM through keyword matching (e.g., [23, 4]) or extraction of related aspects such as consumer needs and opinions through topic modelling [24] and sentiment analysis [25], and then use sentiment as a proxy to revisit intention [26]. However, positive sentiment alone does not necessarily imply positive repurchase intent and vice versa, thus new methods are required to address these limitations by correctly inferring the intention of consumers in the presence of negations as highlighted elsewhere [4]. This is because negations in text can be explicit (e.g., "will not come again") or implicit (e.g., by expressions that invert meanings such as "it is unlikely that

I will come again"). Negation detection has been addressed in other domains that utilise textual data but not in tourism.

Based on the aforementioned gaps in the literature, our first contribution is the use of negation detection to improve classification accuracy of revisit and non-revisit reviews. The second contribution is the extraction of consumers' opinions from eWOM in a bottom up approach through topic modelling and the identification of associations between these opinions and revisit intention using optimised XGBoost classification. This goes beyond the traditional approach of relying on predefined hypotheses justified based on previous knowledge [22]. The third contribution is the application of explainable ML techniques [27] to assist in the interpretation of the generated XGBoost models and the identification of actionable business recommendations and practices for service improvement to enhance revisitation. This paper, thus, provides an empirical analysis of the effects of different inferred factors that describe consumers' perceptions and revisit intention as well as a mechanism for timely business recommendations on how to improve revisit. The paper, therefore, aims to answer the following research questions using and advancing established big data analytic models:

- How can hotel reviews be classified as revisit or non-revisit by analyzing negation in eWOM text?
- What topics discussed by consumers through eWOM can explain tourists revisit intention for two hotel classes (2/3 and 4/5 star hotels)?
- Which topics are most influential on revisit/non revisit in each of the two hotel classes?

To answer these questions the study methodologically combines various techniques from AI and Big data analytics under a unified framework. In particular, Word2Vec language modelling [28] is employed to vectorize eWOM text and identify revisit phrases/words, while the SPACY NLP library is used to filter reviews based on revisit/non revisit intention. A revisit classification model (XGBoost [29]) is trained and then employed to label the primary dataset. Structural topic modelling is used to infer consumer eWOM's opinions, and the topics are then used to train an XGBoost classifier to predict revisit/non-revisit. Finally, the SHAP technique [30] is used to explain inherent patterns of the trained XGBoost model and thus to highlight the most important aspects of consumer revisit/non revisit. At the same time, the combination of negation detection for revisit classification with topic modelling and explainable AI makes this a novel methodology that is generic enough and could be applied on data from different sources to assist businesses manage consumer repurchase in an informed manner.

The paper is organized as follows. "[Literature review](#)" overviews the literature on revisit intention. "[Technical background](#)" provides key technical background on machine learning techniques used for this analysis and "[Research methodology](#)" elaborates on the methodology. "[Results](#)" presents the results from the application of the methodology with data obtained from TripAdvisor. The paper concludes with the discussion, theoretical and practical implications, limitations, and future directions.

Literature review

Consumer revisit intention has been studied for several years in an attempt to understand the patterns that lead to repeated purchase [31]. Such understanding help to establish long-term relationships with customers and facilitate customers' loyalty [32], and to reduce consumers' decision making time and search for alternative products and services [33]. Relationship marketing focuses on customer needs and wants as a way to gain customer loyalty. This requires understanding of consumer motives and in particular, satisfaction [34, 35], which is considered as an essential and direct antecedent of revisit intention [5]. Other influencing factors for tourists' revisit intention include perceived value [36], destination image [37], and memorable experience [38].

Satisfaction [39–41] is defined as a customer's emotional reaction to a product or service experience, and is developed when a customer's perceptions and expectations of service performance are met or exceeded [42]. It is, thus, the difference between the perception after experiencing a service, and the expectation prior to the experience [43]. Tourists with higher satisfaction are more likely to revisit a destination [44]. Thus, satisfaction constitutes a proxy to revisit intention with most studies using satisfaction for inferences about revisit [45]. Antecedents of satisfaction also influence repurchase intentions [5, 46]; therefore, to increase the likelihood of revisit, service providers try to improve features associated with visitors' satisfaction, such as hotel attributes and practices [47, 48]. However, due to different expectations between first time and repeat visitors, hoteliers need to engage these two types of customers differently [48].

In the same vein, loss aversion bias by customers asserts that greater weight is given to losses than gains [49]. The three-factor theory [50, 51] is based on this perspective and examines the asymmetric impact of service/product attributes on overall customer satisfaction. According to the theory, a negative performance on an attribute may create a stronger influence on overall satisfaction than a positive performance on the same attribute [52]. For instance, customers will be dissatisfied if a hotel room is not clean, but they will most likely not be (more) satisfied if the hotel room is clean since this is an expected hotel property/quality. Thus, according to the three-factor theory, satisfaction factors are grouped under the following three categories [50]:

- basic factors are the minimum requirements that cause dissatisfaction if not fulfilled but do not lead to satisfaction if fulfilled or exceeded, implying an asymmetric relationship; e.g., “clean room” since this is considered obvious;
- 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 when present, but do not cause dissatisfaction when absent; e.g., “made feel special”.

Research in revisit intention utilises the three-factor model [48, 53] to assess and explain [54] the impact of prespecified factors on revisit and satisfaction, or uses the popular Theory of Planned Behavior (TPB) [8, 55] and its many extensions (e.g., [56]). TPB studies mainly employ surveys to predict behavioral intention based on constructs, such as attitudes, subjective norms, and perceived behavioral control. Extensions of the method that focus on revisit intention involve constructs such as perceived value

and experience, and essentially use satisfaction as a precursor to revisit [56, 57]. TPB extensions, however, could expand the length of questionnaires that have already been criticised as being time consuming and expensive to run. The main limitation of these methods is the use of prespecified factors that may not apply in novel situations such as the COVID19 pandemic. To overcome this limitation, researchers use interviews to identify important factors to revisit [58]. This method, however, also has limitations in terms of number of participants and time required to collect and analyse the data.

An alternative and comparatively advantageous approach to surveys/questionnaires is the bottom-up analysis that enables the identification of factors that may be causing satisfaction from big data such as eWOM in an unsupervised way. The time required for such analysis is reduced, which in turn enables businesses to react to critical situations in a timely fashion. EWOM is also considered as a more spontaneous source of information in contrast to questionnaires [59]. Consequently, analysis of eWOM from reviews websites has become a mainstream approach for evaluating satisfaction in hospitality and tourism through the use of automated machine learning models [60–62].

However, there is limited research on methods that harness the unstructured part of eWOM for revisit intention analysis with most previous work focusing on the structured part of eWOM such as the rating score (e.g., [63]). Using reviews' rating to infer satisfaction has been criticised as being biased [64] or incorrect, since ratings could be high but the actual review may be negative and vice versa. Hence, when analysing reviews, Valdivia et al. [65] suggest analysing the opinions mentioned in the review in depth instead of using the user rating as a label of sentiment for the whole review. Similarly, Kordzadeh [66] found that there might be biases in star ratings, reducing their reliability. Past evidence [67] 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 indication of satisfaction and thus researchers in tourism have started also using sentiment analysis for reviews satisfaction evaluation [68]. To address these limitations of user ratings, NLP techniques are employed to assess consumers' satisfaction through sentiment analysis on eWOM's text and to use sentiment as a proxy to revisit [26]. However, positive sentiment does not necessarily imply revisit intent, giving rise to the need for NLP-based revisit intention recognition.

Revisit intention is mainly treated as a text classification problem using supervised ML. Thus, there is a need for labelled textual data referring to revisit reviews and non-revisit reviews. Since labelled data are difficult to find or costly to process (annotate) manually [69], researchers seek alternative ways to label text, such as via crowdsourcing. These labels however suffer from noisy data [70]. For instance, Park and colleagues [71] classified re-visitors using the reservation records of a hotel and applied sentiment analysis to find differences among the two visitor groups. Although this study determined revisit intentions, it is impractical since it is difficult to obtain such data from different hotels. Other lines of research seek to harness the structured information of reviews (rating, followers of reviews, liking of review etc.) rather than the textual part, and determine revisit intention using classification or predict consumer's recommending a service using an ensemble of ML techniques. Methods that harness the textual part of reviews use keyword phrases to identify positive and negative intentions for revisit and apply text mining to identify the drivers of each category (e.g., [72]). This approach,

however, could lead to inaccurate conclusions when the true intention of the reviewer as expressed in the text falls outside the scope of the search terms/phrases. Alternative approaches (e.g., [3]) use jointly ML and rule-based classification to infer revisit intention and employ human coders to manually label reviews, which are then used to train the ML models. Manual annotation of large datasets however has two limitations. Firstly, inconsistencies in labels can emerge from annotators with different prior knowledge and experience. This is usually the case with crowdsourcing techniques when multiple labels are generated by different annotators for the same data. Secondly, such process is usually time consuming due to the need to recruit and train annotators for the task and to evaluate the output using methods such as majority voting [73].

Despite the importance of identifying returning customers in different business domains and the abundance of eWOM data, revisit intention recognition using text is not fully exploited with most studies employing keyword matching practices [23, 4] or sentiment analysis through deep learning models such as the work presented in [74], and using the polarity of the text as a proxy for repurchase intention. However, positive sentiment does not necessarily imply positive repurchase intent and vice versa [75]. The association between intention recognition and identification of negation (negative intent) in text is important [76] and more challenging than lexicon-based sentiment analysis and thus requires understanding of the words' context rather than their mere polarity in text [77]. Ignoring negation can impair accuracy when classifying textual information [76]. This is because negation is context-dependent and usually changes the polarity of a sentence, creating an opposition between positive and negative counterparts of the same sentence [78]. To avoid inaccuracies in automated methods for classifying text, it is essential to identify and predict negative text fragments (which may include both negated statements but also those framed positively with negative connotation). This problem has been addressed in the medical domain, and in particular in the analysis of medical reports, where they seek to find negative expressions to rule out certain conditions during diagnosis; for example in the expression "the patient has no fever" the existence of the keyword 'fever' does not denote that the patient has fever [79]. Popular algorithms, such as the NegEx, use rule-based approaches to determine the scope of some negation cues [80]. Alternatively, traditional ML methods tend to ignore information relating to the order of words and their context such as meanings inverted through negations [81]; an example of that is methods based on counting the occurrences of words (or word combinations), resulting in so-called bag-of-words (BOW), which represent text as a matrix with frequencies of words in each document. In addition, ML approaches require more computational resources and manually labelled data, which make them costly and prone to subjective interpretation [3]. Unsupervised ML do not need labelled data, but have worst performance compared to supervised approaches [82]. Despite the importance of negation in text classification, this has not yet been considered in tourist revisit intention research.

Supervised ML models have been successfully used for sentiment analysis; however, when sentiment is used to inform revisit intention, the results are not very accurate since some reviews have positive sentiment but negative intention to revisit [77]. Sentiment analysis' dependence on positive and negative words [83] would make it lean towards revisit rather than non-revisit in the following example "I liked the hotel a lot, but I don't

think I will be coming back because of the staff attitude". Thus, the method proposed in this paper utilizes a combination of rule-based and ML approaches, as suggested elsewhere [3], but goes beyond current methods by developing a novel revisit classification model that handles negation and enables recognition of non-revisit intentions with strong negative connotation. Accurately identifying non-revisit intentions is key to business sustainability since their influence on other consumers decision making is stronger than revisit reviews [21].

Technical background

The core tasks involved in the proposed methodology include text classification, topic modelling and interpretation of the generated ML models. Text classification has as a prerequisite the conversion of text into a machine processable form such as vectors, while topic modelling refers to the process of identifying prevalent topics in a corpus of reviews. The interpretability step of the generated ML models is required when models are complex, with logic and patterns hard to explain to stakeholders. Several techniques exist that can assist ML model explainability. These technical aspects are overviewed below and subsequently it is shown how they are used in the methodology.

Text vectorization techniques

To enable the processing of text using NLP it is essential to either use some metadata (e.g., number of characters, words, nouns/verbs/adverbs, punctuation, etc.) as features/properties of text or convert text in numerical form. In fact, the first task after text cleaning and pre-processing is to convert text in terms of vectors of numbers that act as computational representations of the text. There are two main ways to represent text in numerical form, the traditional feature-based methods and the deep learning methods, both of which are used in this work for different purposes. The deep learning methods represent words in terms of embedding (e.g., Word2Vec) [28], that is, learned in an unsupervised way from large corpus of text to produce word embeddings in the form of multidimensional vector space of continuous values, with similar words having similar vector representations. Word similarity is usually based on neighboring words, so words that appear together have similar representations, while words that rarely appear together would have different embeddings. There are several techniques to achieve word embeddings. In this work, Word2Vec, available through the Gensim library, is adopted due to its popularity and good performance.

An alternative to word embeddings, also used in this study, is the feature-based approaches [84] that break up text into individual words called bag of words (BOW) and treat each word (unigram) or a contiguous sequence of words (2, 3, or more (n) words together referred to as bigrams, trigrams, or n-grams, accordingly) as a potential feature. The splitting of text into words (tokens) is referred to as tokenization. BOW can represent reviews' text as a fixed-length-vector of all terms occurred in the corpus. Each term is weighted either by its frequency or its term-frequency-inverse-document-frequency (TFIDF), which represents the importance of a word to a document by punishing words that are frequent in documents (e.g., is, and, the), hence is able to give more weighting to words that are more relevant and important to a particular document.

Topic modelling techniques

Topic modelling is another text representation method for information extraction from textual data and is used in this work to identify links between the topics discussed in tourists' eWOM and intent to revisit or not the hotel in the future. It belongs to the category of unsupervised data mining techniques employed to reveal and annotate documents with key thematic information [85]. Topic models generally involve a statistical model aiming to find topics that occur in a collection of documents. Two of the most popular techniques for topic analysis are the Latent Dirichlet Allocation [86] and the Structural Topic Model (STM) [87]. The latter is used in this study. During topic modelling, each topic is represented by a set of words that occur frequently together in a corpus and each document by a distribution of topics. Documents in this work are the online tourists' reviews. The process for training the topic model initiates with data pre-processing that includes removal of common and custom stop-words and irrelevant information (punctuation), followed by tokenization and stemming (converting words to their root form).

Subsequently, the optimum number of topics that best fits the dataset is identified through an iterative process examining different values for the number of topics (K) and inspecting the semantic coherence and exclusivity of the model at each iteration until a satisfactory model is produced [87]. Coherence measures the degree of semantic similarity between high scoring words in the topic. Exclusivity measures the extent to which top words in one topic are not top words in other topics.

Text classification process

Text classification, a supervised ML technique that enables the categorization of text into predefined classes, is used in this study to classify reviews as either "revisit" or "not revisit" based on eWOM's text. In contrast to topic modelling, which is an unsupervised ML technique with classes not being predefined, classification requires a labelled dataset (for instance revisit/non revisit reviews). The first task in text classification is to represent text in a format that can be processed computationally since text cannot be fed directly into classification algorithms [22]. With each word or combination of words being a feature, there is a need for feature selection to find the features that have the highest effect on the output variable. Several techniques exist for feature selection including filter methods, wrapper methods, and embedded methods. The filter approaches are independent of the classification algorithm and select features based on their distance (e.g., relief), entropy (information gain) or statistical (e.g., chi-square for categorical, *f*-test for numeric variables) properties. Wrapper and embedded methods use a classifier to evaluate feature subsets as part of their learning process [88]. Embedded methods are integrated into a classifier and thus are faster. In this work, feature selection was performed to (1) identify the features (*n*-grams) that improve revisit classifier performance and (2) find *n*-grams that are associated with sentiment in sentences before and after the target words (during training of the revisit labeler) prior to training the revisit classification model to identify the topics that yield the best model performance.

Traditional text classification methods include Support Vector Machines, Naïve Bayes, Decision Trees (DT), K-Nearest Neighbor, and Tree ensembles models such as Random Forrest and Boosted trees algorithms [82]. Ensemble techniques combine multiple base

classifiers (e.g., DT) in order to achieve a better performance than a single base classifier. Ensemble models compensate for errors in individual models to improve the overall predictive performance. They are preferred for small datasets like in our case, in contrast to Deep models that require large, labelled datasets. Ensembles are also suited for datasets where the features are meaningful (for instance, the features “cleanliness of a hotel” is meaningful) rather than abstract variables that could represent different dimensions of a problem (for instance, pixels in image processing tasks). Bagging and boosting are the two main categories of ensemble techniques. Bagging uses bootstrap sampling (a random sample of data in a training set is selected with replacement) to train independently many base classifiers with each observation having the same probability of being selected, while boosting puts more emphasis on weighting observations and therefore some of them will be sampled more often. Bagging builds a model in a sequential manner considering previous models’ success and adapting weights to misclassified observations [82].

From all types of text classification, binary classification problems are amongst the most popular and refer to the case where there are only two classes for the model to predict, like the “revisit” and “non revisit” in our case. When the number of positive cases (revisit) is much higher than the number of negative (non-revisit), the algorithm is biased towards the positive class and the dataset is termed imbalanced. This is a problem in classification with various metrics used to measure the performance of a classifier on predicting both classes. Several metrics are reported as suitable to evaluate the performance of two-class models, including: (a) sensitivity (b) specificity, (c) overall performance using the area under the receiver operating characteristic curve (AUC), (d) the geometric mean (*G*-mean), and (e) the F-score, the balance between precision (number of True Positives divided by the number of True Positives and False Positives) and recall (number of True Positives divided by the number of True Positives and False Negatives).

The text classification approach used in this work is based on a popular ensemble method, the gradient boosted decision trees, and more specifically its extension the Gradient Boost or XGBoost.

XGBoost [29] has been extensively used in academia and industry due to its good performance and computational efficiency and because it improves traditional boosting through regularisation and optimization of the loss function. Hence, it can deal with model overfitting and produces models that can be generalized [89]. It has been applied in opinion mining tasks and it has become a popular method for finding patterns in eWOM text. XGBoost is an ensemble method, hence multiple trees are constructed with the training of each tree depending on errors from previous trees’ predictions. Gradient descent is used to generate new trees based on all previous trees while optimising for loss and regularisation. The XGBoost regularisation component balances the complexity of the learned model against predictability, while XGBoost optimisation is required to minimise model overfitting and to treat data imbalance, by tuning multiple hyper-parameters.

Interpreting machine learning models

The interpretability of models is key in any area of social/real-life decision-making problems. A model is considered interpretable if it can be visualized or described in plain

language to the end-user [27] or more broadly communicated efficiently to the relevant stakeholders. Interpretability leads to trust in the model while its absence leaves users with little understanding of how particular decisions are made by such models. Black-box models do not disclose any meaningful information about their outputs or their internal structure. Models that are self-explanatory incorporate interpretability directly to their structures. This class of models include decision trees, decision rules, nearest-neighbours, and linear models [27]. Ensembles and deep neural networks are considered black-box models despite their superior predictive performance. Similarly, ensemble methods such as XGBoost, also suffer from limited interpretability that hinders their application in domains where users rely on rational explanations of the models predictions for their decisions [90] or gain domain insights from the inner structure of the model. The lack of interpretability is a substantial obstacle in extracting scientific knowledge from an accurate model.

Different techniques have emerged that tackle interpretability in machine-learning with results categorized into locally and globally interpretable models [91]. The former provide explanations for individual predictions without interpreting the model mechanism as a whole (for instance if you want to know why a particular customer is classified by the model as 'non revisit'). Global interpretability seeks to understand how the model makes decisions, based on a holistic view of its features and each of its learned components such as weights, other parameters, and structures. Notable techniques are the Local Interpretable Model-Agnostic (LIME) [27] and the Shapley Additive explanations (SHAP) [30]. The SHAP method can explain the output of a model through global and local analyses. Local analysis yields a unique Shapley value (denoting importance) for each case or instance, indicating why a case is derived as a specific output. Shapley values can be combined into global explanations.

Research methodology

The aim of this work is the identification of patterns explaining why tourists would like to revisit hotels. To achieve this, two methodological problems are solved. The first involves the labelling of reviews based on the intention of the review author to revisit a hotel or not. To extract all the required information from data, two main NLP techniques are utilised, namely, negation detection and text classification, which are used to predict the category that a review belongs to, for example if a review mentions revisit or non revisit.

The second methodological challenge involves the identification of causes that lead to revisit intention. For this purpose, a different NLP technique is used, namely language topic modelling for finding the topics discussed by tourists in their reviews, followed by two more text classification models for each of two different classes of hotels based on their star rating (2–3 and 4–5 stars). The two classification models are trained from only revisit and non-revisit reviews and use as features the topics discussed in eWOM's text (extracted by the topic modelling) and as class label the previously predicted revisit intention category. The model's patterns are made explicit through the SHAP interpretation technique [30], which helps identify the underlying causes of revisit intentions (or not) using the topics/opinions discussed in eWOM.

Datasets

The methodology is implemented with eWOM data of tourists who stayed and wrote reviews about their accommodation in Cyprus between the years 2010–2019. The data was curated using location filtering criteria in TripAdvisor console and was automatically extracted using a custom-made scraper. TripAdvisor was selected since it is the world's largest travel platform [92] with the most reviews and hotel ratings. Data is publicly available and anonymised, and its use does not involve any privacy or copyright issues.

The total number of reviews collected were 75k, all in the English language, by tourists coming from 27 countries and staying at 2 to 5-star hotels. The timeframe was chosen because of its relative homogeneity in touristic service and intentionally avoided the covid19 period. The data was initially in a comma delimited format and included information about the traveler's username, rating of hotel, name of the hotel, user helpful votes and contributions, dates of stay and date of feedback, city of stay, hotel stars, country of origin, and the review text.

The second dataset utilised in this study is a secondary dataset of 515 K hotel reviews prelabelled by sentiment from booking.com and publicly available [93]. The data refer to reviews from tourists that visited hotels in European countries (not including Cyprus). This data was used to train the revisit intention classifier (labeler), for labelling the primary data (Cyprus hotels reviews) based on the intention of tourists to revisit a hotel or not.

Workflow overview

The workflow of the methodology employed is depicted in Fig. 1. This is composed of the following steps, where the number of each step corresponds to the same-numbered section of the workflow in Fig. 1:

- (A) Data pre-processing: Prior to using the eWOM data for classification, the retrieved data is preprocessed to eliminate emoticons, digits, ascii code, URLs, convert text to lowercase, expand contractions (e.g., “don't” into “do not” etc.) and normalize text by transforming it into a canonical (standard) form. For example, the word “goood” is transformed to “good. The text normalization is performed using a dictionary mapping approach.
- (B) Development of a word embedding model with the pre-processed corpus using the Word2Vec algorithm: The Word2Vec [28] is used to find words similar to the “revisit” term. Tourists' reviews are used to train a 300-dimensions Word2Vec embeddings model, with words with similar meanings clustered together within this space. Word2Vec has two model architectures: continuous bag-of-words or skip-gram. The main difference between them is the input and output data. The skip-gram takes a word as input and predicts its context, whereas continuous bag-of-words takes the context (surrounding words) as input and predicts the missing word. A continuous bag-of-words model is used in this study due to better performance in dealing with common words [94] such as words relating to revisit. A model is trained using data from the reviews and the learned model is used to find

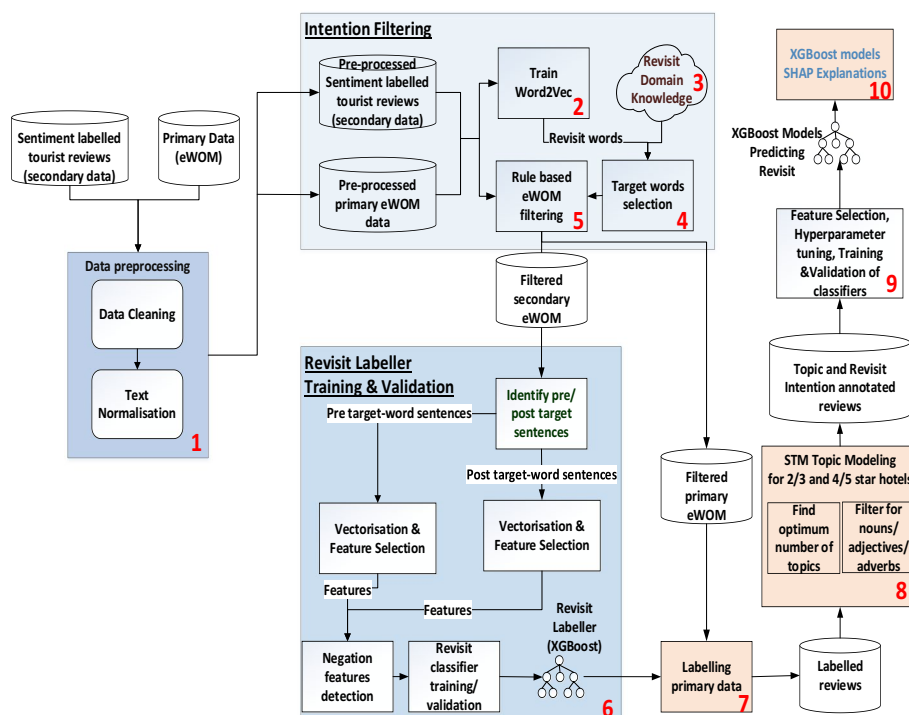


Fig. 1 Overall methodological workflow. Bold numbers refer to steps in methodology

words/phrases relating to revisit intention words. To identify semantically similar words or group of words (n-grams), we use cosine similarity between words’ vectors.

- (C) Identification of terms that refer to revisit intention based on revisit literature (domain knowledge) such as “come back” “revisit”, “stay again”, “return” etc.
- (D) Target words selection: The identified terms are used to find similar words using the trained Word2Vec model. The process is repeated for all identified terms from the literature until the resulting list of terms from the model converges (same terms come up). This step is elaborated further in “[Target words selection](#)”.
- (E) Rule based eWOM filtering: A set of text filtering patterns is created using the identified target words/phrases (e.g., “come back soon”) and both datasets (primary and secondary) are filtered using the specified patterns. This step is elaborated further with a worked example in “[Revisit pattern extraction](#)”.
- (F) Revisit Labeller training and validation: The filtered secondary sentiment-labelled dataset is used to train a revisit classifier (labeller) by utilising the sentences before (pre) and after (post) the identified patterns. Text in the pre/post sentences is vectorised using TFIDF prior to feature selection and training of a revisit labeller. The identification of negation features (unigrams, bigrams) is based on the association of each feature with sentiment. This is elaborated further in “[Revisit labelling](#)”.
- (G) Labelling primary data: The trained “labeller” model is used to label the filtered primary eWOM data with revisit or non-revisit intention.
- (H) Topic Modelling: In this study, the STM approach is used over the LDA due to better topics’ interpretability. The labelled primary data is split into two datasets, one

for the 2/3 star hotels and one for the 4/5 star hotels, and a topic model is built for each hotel category. Initially, common stopwords are considered and gradually additional stopwords irrelevant to our goal are added with the refinement of the model, such as, names of people, hotels, cities, and resorts. In subsequent refinement of the dataset, the corpus is filtered only for verbs, nouns, adjectives, and adverbs that yielded a better model. Tokenization, and in particular the use of phrases composed of n-words (i.e., n-grams) is applied to transform the reviews into a sequence of tokens. The optimum number of topics (K) is identified for each dataset based on coherence and exclusivity metrics. The naming of the topics is performed manually based on domain knowledge and the most prevalent words that characterize each topic. Results related to this step are presented in "[Training the topic models](#)".

- (I) The topics discussed in reviews are used as features to train two XGBoost models to predict revisit/non revisit for each hotel group. The classifiers are optimised by performing feature selection and hyperparameter tuning using Grid-search and manipulating the following hyperparameters: DTs' max_depth, reg_alpha, learning_rate, scale_pos_weight, min_child_weight, and n_estimators. Two text classification models are trained and tested using a 70/30 train/test data split and evaluated against different binary classification metrics such as AUC. This is elaborated in "[Revisit classifier training](#)".
- (J) XGBoost models and SHAP explanations: The SHAP approach is employed to help with the global and local explanation [95] of the learned XGBoost models and thus to help discover patterns that drive revisit/non revisit. Based on these results, stakeholders can make decisions on potential improvements with an aim to enhance revisit performance while also minimizing negative reviews that are damaging to businesses. This is presented in "[Revisit Classifier Interpretation](#)".

Results

The implementation of the methodology yielded results associated with (1) the filtering and labelling of reviews based on revisit intention, (2) the generation of two topic models for the two hotel categories based on their star ratings, (3) the training and validation of two XGBoost classifiers to predict revisit or non revisit intention for the two categories of hotels, and (4) the interpretation of the patterns embedded in the two trained XGBoost classifiers based on which recommendations for the hotel management can be made.

Intention filtering

Since the primary data was not labelled (regarding revisit or non revisit), it was necessary to label the reviews prior to building a classifier to predict revisit and then explain its reasoning. Due to the size of the dataset, manual labelling was not an option and, therefore, a classifier was required to label the data. The first step in building such a classifier is the identification of a secondary labelled dataset. However, such a revisit-intention dataset is not available, hence, a sentiment-labelled dataset with hotel reviews was used, acting as a proxy for revisit intention given that satisfaction is a prerequisite

of revisit intention. The utilization of such labelled data is motivated by our previous work in revisit intention [96] and other similar research literature that uses sentiment as a proxy for revisit [26]. The rationale is that tourists who provide extremely positive reviews will possibly want to revisit a hotel and thus their eWOM might include words pointing to that intention. However, these studies do not explicitly measure intention to revisit, and the reviews selected to draw conclusions are not filtered based on that. Sentiment and revisit intention are highly related but not identical notions; hence, it was essential to first identify reviews that talked about revisit (positively or negatively) prior to classifying their intention. Therefore, reviews were filtered by first identifying target terms related to revisit and intention, and then using these terms to develop a set of pre-specified textual patterns relating to revisit intention.

Target words selection

Revisit-related keywords/phrases (target words) were found from a Word2Vec model trained on the tourists’ eWOM text. The training of the model was performed using bigrams, trigrams, and four-grams phrases from the pre-processed corpus. A similar approach is employed for query expansion in medical information retrieval [97]. Target words identification includes the following steps: terms selection from literature and utilization of the trained Word2Vec model to find similar words for each literature term through a similarity function. The 50 most similar words and phrases for each revisit literature term are collected and subsequently analysed to find the most common words/phrases. These constituted the target terms used subsequently in the analysis. Figure 2 shows an example of the most similar words/phrases to the word “revisit”. This

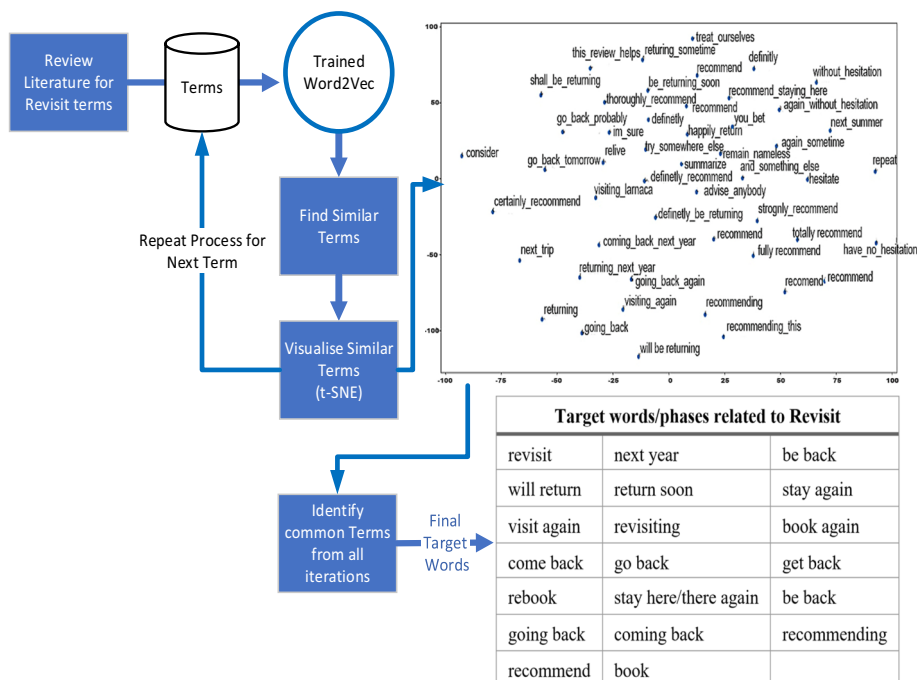


Fig. 2 Utilization of the trained Word2Vec model to find similar terms to revisit terms. The 2D terms plot (top right) shows an example visualization of the top 50 words similar to “revisit” word after t-SNE dimensionality reduction. The table (bottom right) lists the final target words that emerge from this process

is presented in a 2D space after applying a vector space dimensionality reduction technique namely, t-Distributed Stochastic Neighbor Embedding (t-SNE). The identified target terms yielded from this process are depicted in the table of Fig. 2.

Revisit pattern extraction

Reviews were filtered using a set of prespecified textual patterns relating to revisit intention, designed using the identified target terms from the Word2Vec model. Prior to pattern matching, the text was preprocessed, converted to lowercase, and normalized to handle contractions so that these important words for identifying negation in sentences could be used in a cohesive way. The sentiment of the matched reviews was used as the label/class and the words in the pre/post sentences as features to train a revisit classifier. This approach overcomes the challenge of analyzing the entire review text to extract revisit intentions, since reviews can include many intentions about different subjects expressed by the author.

Revisit intention patterns were specified using the SPACY library that offers automated parts-of-speech (POS) recognition for the identification and extraction of generic patterns from text related to revisit. For instance, SPACY patterns can use POS tagging to identify verbs, nouns, adverbs and more, in sentences as illustrated in Fig. 3. Words in boxes refer to the target-term (“come back”) and the words/phrases before (“we will”) and after (“soon to this hotel”) the target term. The example pattern in Fig. 3 is a simplified version of a pattern used in a python script that captures different sentences in reviews that satisfy its conditions. Specifically, the pattern uses lemmas of words (root words, e.g., come and coming have the same root), keywords in sentences that are specific to the task (come, again, soon), and quantifiers to filter sentences based on occurrence of certain POS types of tokens in specific place in a sentence. The quantifiers “*” and “?” specified with the “OP” keyword denote the occurrence of a token one-to-many

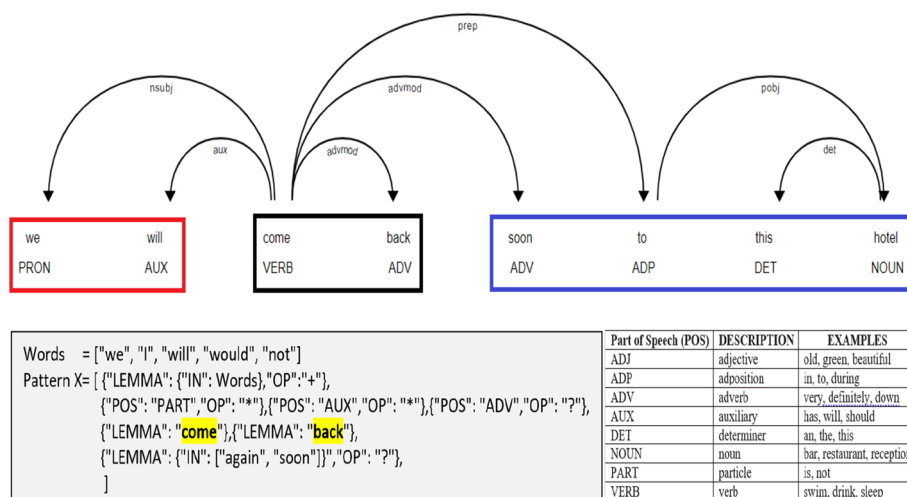


Fig. 3 Sentence parsing with POS annotations (top). The overlaid black rectangle in the middle designates the target phrase “come back” and the red and blue rectangles the pre and post target words, respectively. A SPACY pattern to filter reviews that refer to revisit or non revisit for the “come back” target word (bottom left). A sample of SPACY’s POS relevant to this example (table at bottom right).

times (*) and zero-or-once (?) respectively. The pattern can identify sentences such as: “I will come back soon to this hotel”, “we will not come back to this place”, “we will definitely come back soon to this hotel” and so forth. The benefit of this approach is that many variations of intention sentences can be filtered with a single generic pattern. In this work, 15 patterns were specified to capture all possible intention sentences based on target terms identified from Word2Vec. The example pattern in Fig. 3 and its POS parsing illustrate how SPACY can identify sentences in a review’s text that satisfy its rules. We use lemmatized words so that the pattern can equally detect both “coming back” and “come back” with the same root word (i.e., “come”). The pattern X specifies that for a sentence to be a match, the lemmas of its first words must be in the set of prespecified set of words {Words}. These words must occur at least once in the sentence (use of “+” quantifier) and can be followed by zero to many (use of “*” quantifier) POS {PART, AUX} (e.g., not) and zero or once (use of the “?” quantifier) POS{ADV} (e.g., definitely, surely etc.). All previous words and POS must be followed by the lemmas of the words {come}{back}. All previous words can be followed, or not (zero to one time, use of the “?” quantifier), by the lemmas of the words {again} or {soon} and so forth. Such patterns enable the identification of revisit-intents that can be expressed in different ways without explicitly specifying the sentences of interest. Thus, the example SPACY pattern in Fig. 3 can identify sentences in text such as “I will come back soon”, or the negated sentence “we will not come back again”.

Revisit labelling

The classification of revisit intent is based on tokens of one sentence before and one after the identified revisit text patterns in reviews (pre/post target sentences). Words before and after the target terms (Table in Fig. 2) within the pattern are used as features to train two classifiers using as class/label the known sentiment of the review. In this way, the revisit classification task was not blindly based on the sentiment of the review but also on the presence of the revisit pattern. To address the problem of negated intent, two classifiers were trained: one with the words before and the other with the words after the target terms. This was essential since negation can be expressed by tokens before but also after the target terms. For instance, the simple example with the negation before the target term is “I would not come again to this hotel” while a more complex case is when the negation is after the target term such as “would I revisit? I do not think so”. To address possible false negation by the pre-target classifier, n-grams were used to identify word patterns such as in the example “I can not wait to come again” with the 3-gram “can-not-wait” being associated with revisit rather than non-revisit by the classifier since reviews with such pattern were labelled with positive sentiment, while the pre-target bigram feature “would-not” pushes the classifier prediction towards the non-revisit prediction due to the association of this bigram with negative sentiment label. The features from both classifiers were used collectively to train the revisit labeler.

Text vectorization and feature selection from secondary data

TFIDF and n-grams were used to vectorize the text in sentences before and after the target words in the secondary dataset that was sentiment-labelled. Two TFIDF matrices were produced representing terms (n-grams) influential to negative and positive

sentiments before and after the target terms. These constituted the initial features of the XGBoost labeler. To reduce the dimensionality of these matrices, two feature selection methods were used, namely embedded methods using random forest and correlation-based filtered methods. Wrapper methods were not considered in this study since they are computationally expensive and sensitive to the classifier. The embedded method utilised in this work used the Random Forest model. The correlation-based approach ranked features based on the correlation with the target variable and no relationship to any other features in the dataset. Highly correlated features in both pre and post sentences were eliminated. Further feature selection was achieved using a random forest feature selection embedded method, optimized on the AUC score using recursive feature elimination with cross-validation (RFECV), as depicted in Fig. 4. The resulting features after feature selection were 441 for post target and 641 for pre target sentences.

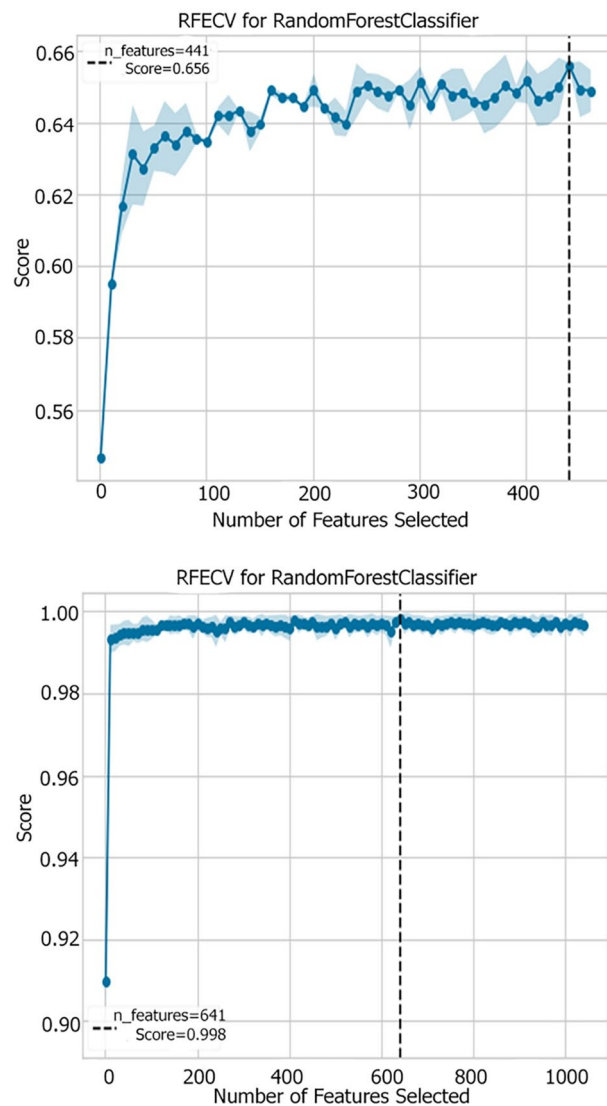


Fig. 4 Featureselection using RFECV and random forest, for words (features) before (bottom)and after (top) the target terms

An XGBoost classifier was trained using as features the combination of the selected pre and post target features and the sentiment as class label. The training was performed using train/test (70/30) data split and was optimized on classifiers' AUC metric. The trained classifier was used to label revisit intention in all reviews. The classifier's optimum classification threshold was identified from the AUC scores. Thus, instead of classifying reviews using the naive approach which uses cases with probability greater than 0.5 as revisit and less as non revisit, the labelling was made based on the threshold values that yielded the best AUC. In this way, the confusion matrix, which is a summary of prediction results on a classification problem, yielded the best possible outcomes.

Revisit intention descriptive results

Overall, from the 75 K reviews, 17 K explicitly mentioned the author's intention to revisit or not. Descriptive results of these reviews, depicted in Fig. 5, show the probability of expressing revisit intentions across 10 years over all reviews per year, including reviews with no explicit reference to revisit/no revisit (not shown in Fig. 5). The confidence interval around the lines (estimated central tendency) shows the variability in the probability for each category of predictions. Therefore, the calculated probabilities are 'revisit', 'non revisit' and 'not explicitly stating any revisit/non revisit intent', all three totaling to 1. Since we are interested in revisit/non-revisit, we depict the probabilities of only these two cases of explicit mentioning of either revisit/non revisit in Fig. 5.

It can be observed that 2/3 star hotels have a higher overall revisit probability than 4/5 hotels. However, there is a stronger negative trend in revisit and non-revisit intentions for 2/3 stars between 2017 and onwards in contrast to 4/5 star hotels. This reduced eWOM with explicit reference to revisit or non revisit maybe due to hotels satisfying the bare minimum of consumer needs, i.e., basic factors (three factor theory). This eWOM behavior could be attributed to the impact of the economic crisis in Cyprus with reduced investments by hotels. Additionally, the probability variability for 2/3 star hotels in comparison to 4/5 star is higher, which could be attributed to higher homogeneity in the quality of service and amenities offered by 4/5 star compared to the 2/3 star hotels. Therefore, tourists have higher chances of meeting their expectation in 4/5 hotels, since these hotels' features are more scrutinized than those of 2/3 star hotels.

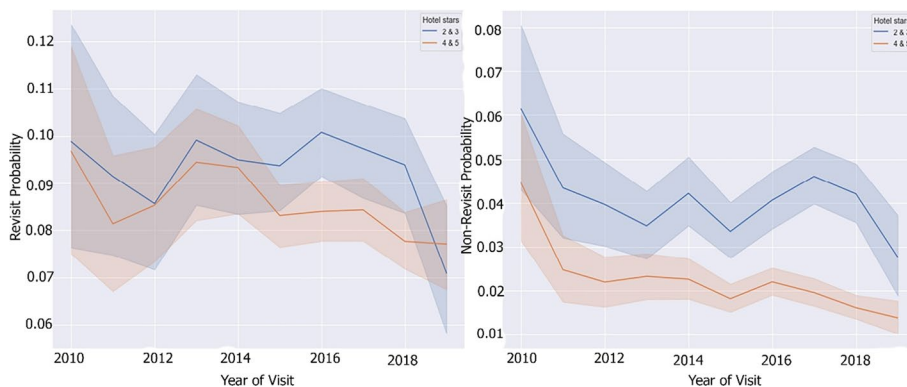


Fig. 5 Probability and confidence interval of revisit and non-revisit across the years 2010–2019

Training the topic models

Two STM topic models were developed (one for each hotel category) using only reviews that explicitly stated intention to revisit or not ($N = 17\text{ K}$). The learned topic model is expressed as a probability distribution of topics per review and denotes the probability of each topic discussed in a review; all topics' probabilities in each review sum up to 1. The trained STM models' theta values that refer to the probability that a topic is associated with each review are used as features during the training of the text classifiers.

Prior to topic learning, reviews had to be pre-processed further to eliminate irrelevant information through stop-word removal, stemming, and tokenization. This preprocessing differs from the intent recognition step that did not include elimination of stop-words (for example "not", since such words are important for negation detection) and words less than 3 characters, and stemming. Two approaches were employed prior to finding the optimal number of topics (k) for each corpus (2/3 and 4/5 star hotels). The first approach used all preprocessed text and the second only nouns, verbs, adjectives, and adverbs. The datasets for each category of hotels were used to build two different models using different numbers of k (10–50): these were evaluated in terms of exclusivity and coherence [87] to find the optimum value for k as depicted in Fig. 6. Based on the results and relevant recommendations [87], we consider 46 topics for the 2/3 star hotels' topic model and 43 for the 4/5 topic model. These analyses were conducted with the STM package [98] in R.

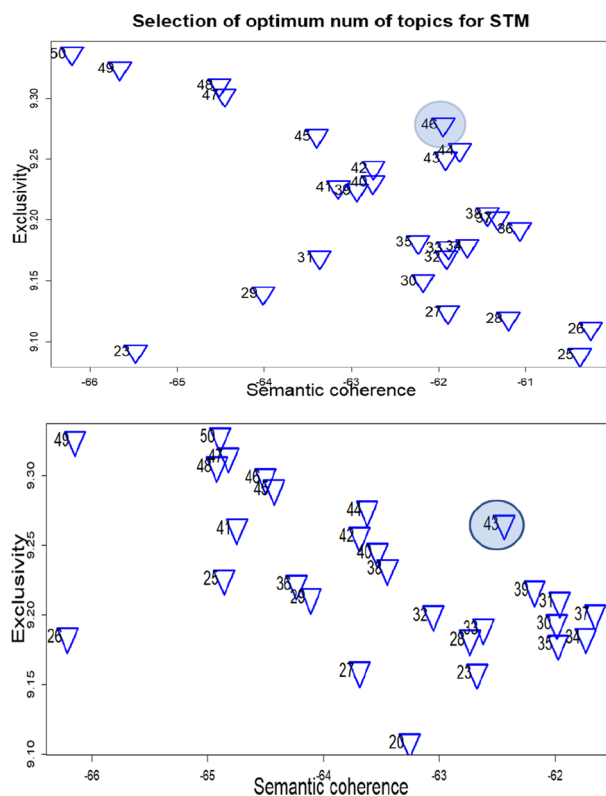


Fig. 6 Semantic coherence and exclusivity of 50 topic models analysed to identify the best k for the 2/3 star (top) and 4/5 star (bottom) hotels models. The optimum number of k is circled for each case

For the interpretation of topics, we first considered words highly associated with each topic, then inspected the most prevalent reviews related to each topic, and finally mapped each topic with prevalent hotel service quality factors. Based on the hospitality literature [63, 99, 100, 101], factors are categorized into tangible and intangible; the former include cleanliness, location, room amenities, quality of mattress, entertainment, lighting and hotel facilities such as Wi-Fi. On the other hand, intangible attributes focus on hotel atmosphere, employee friendliness, and service quality such as reliability (e.g., punctuality), responsiveness (e.g., prompt service), assurance (e.g., politeness), and empathy (e.g., personal attention).

Revisit classifier training

The datasets that emerged from the labelled reviews and the topics associated with each review from the two categories of hotels were used to train two XGBoost classification models that predicted intention to revisit. The topics' probability distributions per review were used as the input features of the model. Prior to training the two models, feature selection was performed to identify the topics that would yield the best classification performance using a random forest and recursive feature elimination with cross-validation. As depicted in Fig. 7 and 40 features (topics) were considered while building the XGBoost classifier for the 2/3 star hotels, and 31 features for the 4/5 star hotels classifier.

To account for the data imbalance (number of revisits much higher than non-revisit), and to maximize the efficiency of the model, the grid search approach was used to select the best hyperparameters for achieving the best AUC classification metric by considering all combinations of hyperparameters that mitigate data imbalance. Figure 8 shows the improvement in 2/3 star XGBoost model's performance after hyperparameter tuning. To train and test the models, the data was split into training/testing sets using stratified train-test split to preserve the same proportions of revisit/non revisit cases as in the original dataset. The two learned models yielded a prediction accuracy of 78% in the test data, 80% F-score as an indication of precision and recall measurement, and 83% effective class separation measured using the AUC for the 2/3 star hotels and respectively, 79%, 82% and 85% for the 4/5 star hotels.

Revisit classifier interpretation

Using the SHAP technique, the XGBoost models' inherent patterns were externalized. SHAP assigns each feature of the model (topics) with an importance value, also known as Shapley value, used to estimate the importance and effect of each feature on the model's output [102] and the interaction between variables [30].

The SHAP summary plots of Fig. 9 combine each model's feature importance with their effects on the target variable (revisit) in terms of log-odds. Each point on the summary plot is a Shapley value for an instance of a feature (for example the topic "rooms-poor etc."). The features are ordered according to their importance on the vertical axis. The color of the points represents the intensity of the topic discussion in a review (feature) from low (blue) to high (red) intensity. The points on the graph form a distribution of the Shapley values per feature.

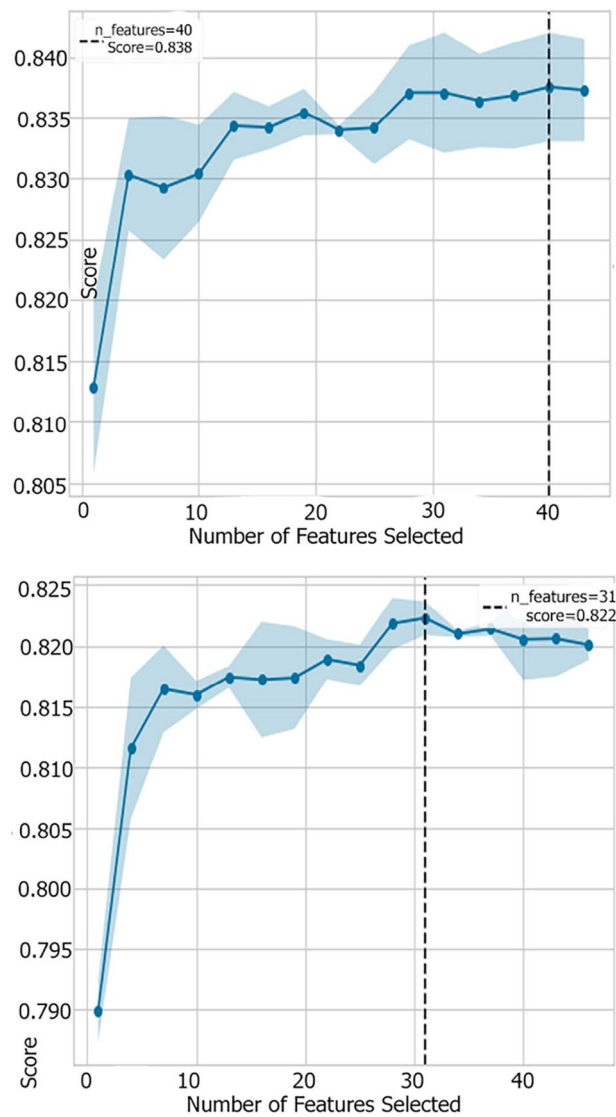


Fig. 7 Feature selection (RFECV) prior to training the XGBoost model for the 2/3 star hotels (top) and 4/5 star hotels (bottom)

The summary plots show the association between topics and revisit. Dots on the SHAP summary plot represent single observations (i.e. reviews). The horizontal axis refers to the SHAP value which denotes the average marginal contribution of the feature value to the output across all possible coalitions (in game theoretic terms). Negative SHAP values (below zero) indicate a negative contribution (i.e., non-revisit), equal to zero indicates no contribution, and positive values indicate a positive contribution (i.e., revisit). For example, in Fig. 9 the probability of a customer non-revisiting is high if the red dots are placed on the left-hand side of the vertical line (negative SHAP values) for each of the features and vice versa.

The SHAP summary plots (Fig. 9) highlight important variables as these are extracted from the topic models that affect revisit intention. For the 4/5 star hotels the most important topics that affect revisit are the rooms and bathroom age/state (Fig. 9 bottom),

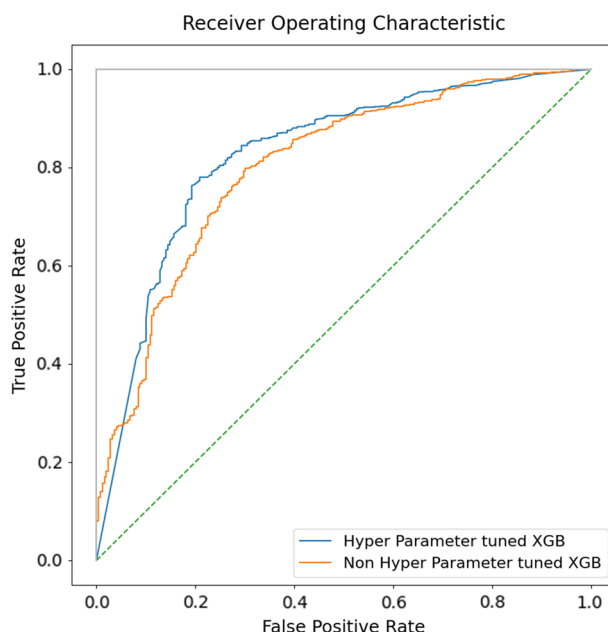


Fig. 8 XGBoostclassifier’s AUC (2/3 star hotels) before (orange line) and after (blue line) hyperparametertuning

with more discussion about the quality and age of the bathroom negatively affecting the intention of the tourist to revisit. Second topic in the ranking is the cleanliness of the rooms and the quality of the food. When both aspects are positively discussed by tourists, then the intention to revisit is increased. Additional negative topics are arguments with the reception regarding various issues, the requirement to pay for drinks in all-inclusive packages, noise at night, and quality of food. On the contrary, aspects that improve revisit probability are the service quality and cleanliness. Overall, the Shapley values of negative topics on revisit are greater than the Shapley values of positive factors on revisit, which confirms the three-factor theory that highlights the asymmetric impact of basic service/product attributes on overall customer value perception (three factor theory).

The SHAP summary plot for 2/3 star hotels as depicted in Fig. 9 highlights the following topics as positively contributing to revisit: comfort and location of the accommodation, staff, breakfast and cleanliness. On the contrary, negatively contributing topics include, poor room quality, dirty bathroom, entertainment, air conditioner quality during sleeping, inclusive drinks.

In addition to the summary plots, the dependency plots in Fig. 10 provide a drill down topic analysis by showing the Shapley values in accordance to revisit per individual topic. These plots show that as the intensity of the topic discussions(x axis) increases in eWOM, the effect on revisit is stronger. The impact of the effect can be obtained from the shift in the log odds on the y-axis while the tolerance of the tourist to the issue discussed in the topic is denoted by the angle of the local weighted regression line that describes the trend. For instance, the tolerance of tourists to the topic ‘old-room-bathroom’ (simplified topic name based on words associated with it) is less than that of ‘pay-inclusive-drinks’ which again goes back to the asymmetric

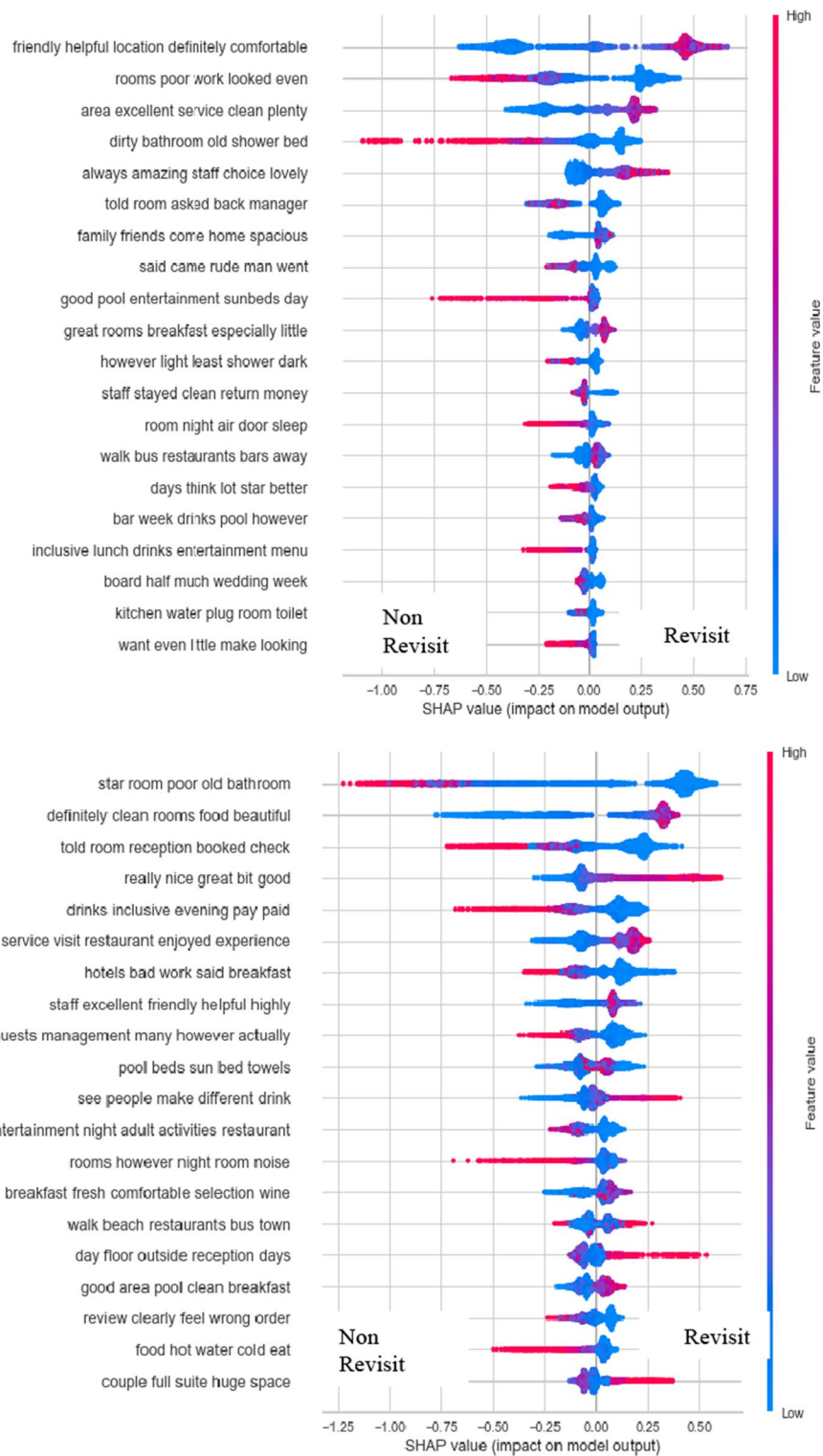


Fig. 9 SHAP summary plots for the 2/3star (top) and 4/5 star (bottom) hotels showing most prevalent words in each topic as features

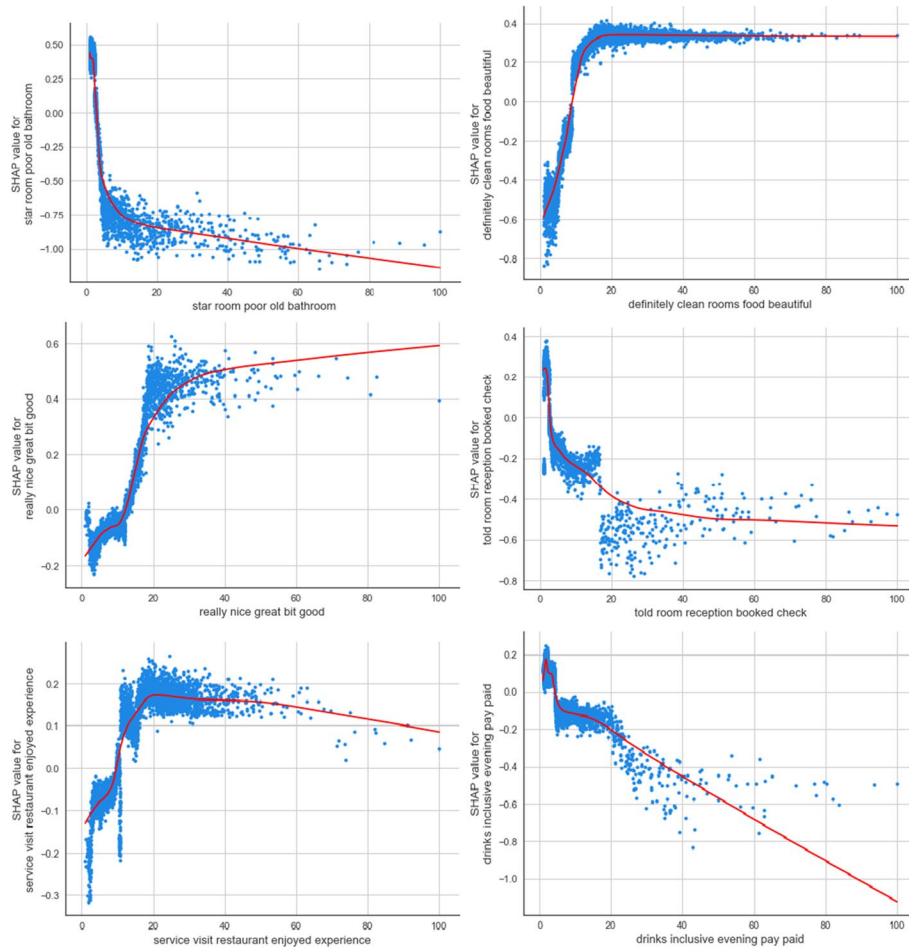


Fig. 10 SHAPdependency plots of the 6 most important topics (4/5 star hotels). The y axis refers to revisit probability expressed as log odds and x axis is the topic intensity in reviews. The red line is the trend

relationship between positive and negative effects of basic factors (three factor theory). Similarly, the quality of the food and the cleanliness have greater impact on revisit compared to other topics. The identification of such basic factors is key to improve consumer repurchase since revisit performance will not improve by focusing on excitement factors when the basic factors are not there.

It is evident from the results that the topics that cover the two groups of hotels have some common themes such as age of the facilities and the all-inclusive deal. However, the 2/3 star hotels also have additional negative topics such as cleanliness of the bathroom, pool entertainment, sunbeds availability, and air conditioner quality, while the 4/5 stars have issues with noise and guest management.

These results constitute recommendations for the hotel management on how to improve service to maximise revisit. Thus, the method showcases a timely approach for dealing with revisit given the many benefits of revisitors to businesses and the negative effects of explicit non revisit intention reference in eWOM.

Discussion of results

This work provides a novel method for investigating revisit intention and extends previous methods (e.g., [17, 26]) of eWOM sentiment to infer about revisit, satisfaction [103] and quality of service from big data in the hospitality industry [104]. The method integrates negation detection, topic modelling, and SHAP-based interpretation of XGBoost classifiers and is similar to earlier work using eWOM to extract revisit intention with a combination of rule-based and ML techniques [3] and using ML to classify eWOM to harness a lexicon for non-revisitors and revisitors [26]. However, in contrast to the aforementioned earlier studies [26, 3], our work expands on models' explainability and performance improvement through the use of negation detection without the need for manual labelling of data. A similar negation detection approach, but with manually labelled data, was used to assess sentiment and generate summaries of consumers' eWOM [68].

Two XGBoost models are trained herein using as features the topics discussed in tourists' reviews and as predictor the label allocated to the reviews after revisit intention annotation (as revisit or non revisit). The models were shown to be effective in predicting revisit/not-revisit intention with AUC greater than 83% in both hotel categories, and thus in identifying patterns that may lead to these intentions. Topics discussed in reviews are used to understand tourists' positive/negative revisit intention; consequently, the factors identified are not specified in advance, but rather emerge from the data. In contrast to other revisit analysis studies that either use manually annotated data [3] or utilise review sentiment or rating as a proxy to revisit [26], the proposed method uses ML annotation and thus enable faster execution and generation of results.

The study uncovers and assesses factors that are beyond a set of predefined criteria explored in survey-based research [56, 57] and does that in an automated manner, thus enabling businesses to react to critical situations in a timely manner. Similar to Xu et al. [105], the study applies the three factor theory to understand the symmetric and asymmetric effects of eWOM' topics on revisit. Some of the factors identified herein are confirmed as important in previous research such as basic factors (according to the three factor theory) relating to cleanliness and service quality [106, 63, 99, 100, 101]. Several other factors are also identified in this study including the positive but also negative effect of "entertainment" (similar to [48]) and the negative effect of "all inclusive" schemes that do not cover the cost of all beverages. This study identifies both negative and positive factors affecting customer value perception.

The identification of negative factors is important with more significant practical implications, since these have higher impact on revisit performance as identified in the results and highlighted in the literature. For example, "rude attitude of employees", one of the important negative factors, is also highlighted elsewhere [107]. Such negative factors can be further scrutinized using the dependency SHAP plots and in combination with positive factors enable the investigation of revisit in a holistic manner providing a more in-depth interpretation of consumer behavior.

Conclusions

This research addresses the problem of hotel revisits management through the analysis of eWOM content. The proposed method integrates negation detection, topic modelling, and text classification techniques to improve the accuracy of the produced results.

The approach enables the automated analysis of revisit intention from eWOM as well as the provision of timely results. Traditional approaches to revisit analysis use surveys or interviews; however, these techniques are expensive, time consuming, and cannot be always automated. Our approach utilized free online information from the web and hence has a comparative monetary advantage to these approaches.

Theoretical implications

The proposed method has theoretical and methodological implications for various fields. Considering the limited research in the literature that employs ML and NLP to investigate revisit intention in the hospitality industry, the proposed methodology can be useful for future studies as a basis for new proposals for knowledge creation and extraction of insights from other social networks such as Twitter or Yelp.

Researchers can apply the proposed method on eWOM from different consumer segments, such as first-time visitors or loyal consumers, tourists from different cultural backgrounds, economic conditions, or different behavioural attitudes (novelty seeking etc.), to inspect, assess, and understand what aspects these customers value most when stating that they are willing to revisit or not. Negative factors that reduce revisit intention in these consumer segments can also be analysed using this approach. The proposed approach can also be used to validate insights extracted with traditional means regarding the relationships between revisit with other variables.

The methodology can also be applied in different contexts, other than hotel revisit, such as restaurant and destination re-visit, or product repurchase using product reviews. The proposed methodology also addresses shortcomings in measuring revisit intention through traditional means such as surveys with questionnaires, since reviews are more spontaneous and are generated voluntarily based on customers' experiences. Hence, extracting information from eWOM could avoid the self-reporting bias that usually occurs with questionnaires. Furthermore, the topics/factors identified in the present study can be explored as variables using quantitative models that can explore more generalizable inferential models. Thus, scholars can use the present investigation as a precursor of future quantitative investigations through the formulation of hypotheses in empirical studies. Scholars may construct statistical models and justify theoretical variables related to revisit using questionnaires to either verify the results or scrutinize the variables further.

Practical implications

The revisit intention probability calculated by the trained classification model could be used as a measure of hotels' performance. For instance, such metric can illustrate how good the products and services of hotels are as opposed to the currently used review ratings, to monitor customers' satisfaction. This metric could provide alternative means to evaluate how well a hotel performs in terms of earning customers' loyalty. In the absence of existing revisit intention labels in reviews, hotels can use the proposed approach to assess their revisit performance automatically instead of manually processing reviews or conducting timely surveys or interviews. Similarly, destinations' management can use the revisit intention metrics and diagnostics to shape their policy and marketing strategy. Such methods can be incorporated in customer relationship management tools to

assist management in attaining strategic goals with regards to retainment of consumers [108].

The method enables the automatic identification and assessment of the impact of “performance”, “excitement” and “basic” factors (three factor theory) on revisit intention. Since positive revisit intention relates to what hotels do well (excitement and performance factors), while non revisit intention highlights what needs to be changed (basic factors not met), the probabilities associated with each factor can be used to prioritize what hoteliers need to focus on first to have the highest impact on their revisit performance in the less possible time. Hotels can use the proposed method on all available textual information written about them in different social media platforms by consumers to identify the adjustments needed for their operations to improve customers’ revisit. Moreover, since the expectations between first time and repeating customers differ [48], hoteliers need to engage with each of these customer groups differently. Insights from the application of the method can assist hoteliers to formulate the most appropriate strategies for each group.

Non revisit intention is usually accompanied with negative eWOM, thus if not addressed early, it could have serious consequences on hotel’s reputation and profitability. Hoteliers need to learn from the feedback on poorly offered services and react quickly by improving their operations to avoid additional negative impacts. The automated approach proposed in this study can help in the speedy response to these challenges. Repeating customers on the other hand intend to spread positive eWOM and thus could increase hotels’ bookings, therefore, the drivers of revisit need to be identified and utilised wisely through advertisements to attract more customers.

The revisit performance metrics can also be used by social media platforms such as TripAdvisor to provide customers with additional means to support their decision making. High revisit performance of a hotel represents a higher chance of a customer having a positive experience. Such information in combination with the extracted factors that contribute to revisit performance could have a greater influence on consumers purchasing decision compared to review ratings or the manual evaluation of reviews’ text. Therefore, such functionality could attract more traffic (and as a result more profits) to such websites because it would reduce customers’ choice uncertainty.

Limitations and future research

The work presented comes with some limitations. Firstly, the proposed model overlooks certain factors that could influence revisit intention, such as brand name or contextual variables such as the weather [109]. These factors can be used as moderators in future studies.

Secondly, this work does not address fake reviews and does not consider the credibility of the eWOM author as also noted elsewhere [13]. Such information could influence the results. Fake reviews can be addressed through a different classifier and reviewer’s credibility can be harnessed from the followers and likes. Moreover, from the classification performance perspective, the revisit classifier can be further scrutinized by comparing its performance against classifiers trained on crowdsourced data of consumer revisit intention, collected through an experiment on Amazon’s Mechanical Turk and validated

using majority vote [73] since it is the simplest, yet most effective, ground truth inference algorithm for crowdsourced data.

Further work will evaluate an additional explainable machine learning method to assist in identifying the changes that businesses need to perform to their practices to guarantee increase in revisits using counterfactual explanations of machine learning models. Counterfactual explanation methods [110] have attracted increasing attention in recent years since they can be used to explain and recommend actions to be performed to obtain the desired outcome. Future work is directed towards using counterfactual explanations, focusing on multiple non-revisit cases, to identify the minimum interventions that an organisation needs to make to its practices/policies to guarantee increased revisitors.

Finally, our future work will also seek to exploit deep neural networks for intent classification, such as Bidirectional Encoder Representations from Transformers (BERT) [111] that have been recently applied in other text classification tasks with success. However, these models are computationally intensive, which limits their application to GPU-based processing. In the future, we will utilise BERT to benchmark the results to our approach.

Acknowledgements

Not applicable.

Author contributions

AG: conceptualization, methodology, designed the models, validation, software development, performed analysis, interpretation of results, visualisations, writing and editing, supervision. MP: conceptualization, writing and editing. HH: methodology, writing and editing. EC: software development, data curation, designed the models. All the authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 7 July 2022 Accepted: 20 April 2023

Published online: 06 May 2023

References

1. Lee J, Lee H, Chung N. The impact of customers' prior online experience on future hotel usage behavior. *Int J Hosp Manag.* 2020;91:102669.
2. Sun Q, Dong M, Tan A. An order allocation methodology based on customer repurchase motivation drivers using blockchain technology. *Electron Commer Res Appl.* 2022;56:101218.
3. Liu Y, Beldona S. Extracting revisit intentions from social media big data: a rule-based classification model. *Int J Contemp Hosp Manag.* 2021;33(6):2176–93.
4. Ravula P, Jha S, Biswas A. Relative persuasiveness of repurchase intentions versus recommendations in online reviews. *J Retail.* 2022;98(4):724–40.
5. Um S, Chon K, Ro Y. Antecedents of revisit intention. *Ann Tour Res.* 2006;33(4):1141–58.
6. Yu J, Seo J, Hyun SS. Perceived hygiene attributes in the hotel industry: customer retention amid the COVID-19 crisis. *Int J Hos Manag.* 2021. <https://doi.org/10.1016/j.ijhm.2020.102768>.
7. Yu M, Cheng M, Yang L, Yu Z. Hotel guest satisfaction during COVID-19 outbreak: the moderating role of crisis response strategy. *Tour Manag.* 2022;93:104618.

8. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. 1991;50(2):179–211.
9. Huang S, Hsu CHC. Effects of travel motivation, past experience, perceived constraint, and attitude on revisit intention. *J Travel Res*. 2009;48(1):29–44.
10. Nazir S, Khadim S, Ali Asadullah M, Syed N. Exploring the influence of artificial intelligence technology on consumer repurchase intention: the mediation and moderation approach. *Technol Soc*. 2023;72:102190.
11. Tajeddini K, Mostafa Rasoolimanesh S, Chathurika Gamage T, Martin E. Exploring the visitors' decision-making process for Airbnb and hotel accommodations using value-attitude-behavior and theory of planned behavior. *Int J Hosp Manag*. 2021;96:102950.
12. Wei J, et al. The impact of negative emotions and relationship quality on consumers' repurchase intention: an empirical study based on service recovery in China's online travel agencies. *Heliyon*. 2023;9(1):e12919.
13. Verma D, Dewani PP, Behl A, Dwivedi YK. Understanding the impact of eWOM communication through the lens of information adoption model: a meta-analytic structural equation modeling perspective. *Comput Hum Behav*. 2023;143:107710.
14. Serra Cantallops A, Salvi F. New consumer behavior: a review of research on eWOM and hotels. *Int J Hosp Manag*. 2014;36:41–51.
15. Mauri AG, Minazzi R. Web reviews influence on expectations and purchasing intentions of hotel potential customers. *Int J Hosp Manag*. 2013;34:99–107.
16. Sparks BA, Browning V. The impact of online reviews on hotel booking intentions and perception of trust. *Tour Manag*. 2011;32(6):1310–23.
17. Mehra P. Unexpected surprise: emotion analysis and aspect based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists. *Tour Manag Perspect*. 2023;45:101063.
18. Sotiriadis MD, van Zyl C. Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists. *Electron Commer Res*. 2013;13(1):103–24.
19. Gerdt S-O, Wagner E, Schewe G. The relationship between sustainability and customer satisfaction in hospitality: an explorative investigation using eWOM as a data source. *Tour Manag*. 2019;74:155–72.
20. Hu F, Teichert T, Liu Y, Li H, Gundyeva E. Evolving customer expectations of hospitality services: differences in attribute effects on satisfaction and re-patronage. *Tour Manag*. 2019;74:345–57.
21. Zhu JJ, Chang Y-C, Ku C-H, Li SY, Chen C-J. Online critical review classification in response strategy and service provider rating: algorithms from heuristic processing, sentiment analysis to deep learning. *J Bus Res*. 2021;129:860–77.
22. Xiang Z, Schwartz Z, Gerdes JH, Uysal M. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int J Hospitality Manage*. 2015;44:120–30.
23. Zhang N, Liu R, Zhang X-Y, Pang Z-L. The impact of consumer perceived value on repeat purchase intention based on online reviews: by the method of text mining. *Data Sci Manag*. 2021;3:22–32.
24. Subroto A, Christianis M. Rating prediction of peer-to-peer accommodation through attributes and topics from customer review. *J Big Data*. 2021;8(1):9.
25. Feizollah A, Mostafa MM, Sulaiman A, Zakaria Z, Firdaus A. Exploring halal tourism tweets on social media. *J Big Data*. 2021;8(1):72.
26. Chang J-R, Chen M-Y, Chen L-S, Tseng S-C. Why customers don't revisit in Tourism and Hospitality Industry? *IEEE Access*. 2019;7:146588–606.
27. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You? Explaining the predictions of any classifier," NAACL-HLT 2016–2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Demonstr. Sess, pp. 97–101, 2016.
28. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv*. 2013. <https://doi.org/10.48550/arXiv.1301.3781>.
29. Chen T, Guestrin C. "XGBoost: A scalable tree boosting system," in Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
30. Lundberg S, Lee S-I. "A Unified Approach to Interpreting Model Predictions," *Adv. Neural Inf. Process. Syst*, pp. 4768–4777, 2017. <http://arxiv.org/abs/1705.07874>.
31. Jang S, Feng R. Temporal destination revisit intention: the effects of novelty seeking and satisfaction. *Tour Manag*. 2007;28(2):580–90.
32. Ndubisi NO. Relationship marketing and customer loyalty. *Mark Intell Plan*. 2007;25(1):98–106.
33. Jane Hollowell C, Rowland Z, Klietstik T, Klietstikova J, Dengov VV. Customer loyalty in the sharing economy platforms: how digital personal reputation and feedback systems facilitate interaction and trust between strangers. *J Self-Governance Manag Econ*. 2019;7(1):13–8.
34. Bandyopadhyay S, Martell M. Does attitudinal loyalty influence behavioral loyalty? A theoretical and empirical study. *J Retail Consum Serv*. 2007;14(1):35–44.
35. Dick AS, Basu K. Customer loyalty: toward an integrated conceptual framework. *J Acad Mark Sci*. 1994;22(2):99–113.
36. Petrick JF, Backman SJ. An examination of the construct of perceived value for the prediction of golf travelers' intentions to revisit. *J Travel Res*. 2002;41(1):38–45.
37. Chew EYT, Jahari SA. Destination image as a mediator between perceived risks and revisit intention: a case of post-disaster Japan. *Tour Manag*. 2014;40:382–93.
38. Kim JH. The impact of memorable tourism experiences on loyalty behaviors: the mediating effects of destination image and satisfaction. *J Travel Res*. 2018;57(7):856–70.
39. Hultman M, Papadopoulou C, Oghazi P, Opoku R. Branding the hotel industry: the effect of step-up versus step-down brand extensions. *J Bus Res*. 2021;124:560–70.
40. Assaker G, O'Connor P, El-Haddad R. Examining an integrated model of green image, perceived quality, satisfaction, trust, and loyalty in upscale hotels. *J Hosp Mark Manag*. 2020;29(8):934–55.
41. Hu H-H, Kandampully J, Juwaheer TD. Relationships and impacts of service quality, perceived value, customer satisfaction, and image: an empirical study. *Serv Ind J*. 2009;29(2):111–25.
42. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. *J Mark Res*. 1980;17(4):460.

43. Sultan P, Wong hoY. Service quality in higher education – a review and research agenda. *Int J Qual Serv Sci*. 2010;2(2):259–72.
44. Seetanah B, Teeroovengadam V, Nunkoo R. Destination satisfaction and revisit intention of tourists: does the quality of Airport Services Matter? *J Hosp Tour Res*. 2020;44(1):134–48.
45. Alaei AR, Becken S, Stantic B. Sentiment analysis in tourism: capitalizing on Big Data. *J Travel Res*. 2019;58(2):175–91.
46. Ali F, Ryu K, Hussain K. Influence of experiences on memories, satisfaction and behavioral intentions: a study of creative tourism. *J Travel Tour Mark*. 2016;33(1):85–100.
47. Kim WG, Li JJ, Han JS, Kim Y. The influence of recent hotel amenities and green practices on guests' price premium and revisit intention. *Tour Econ*. 2017;23(3):577–93.
48. Lai IKW, Hitchcock M. Sources of satisfaction with luxury hotels for new, repeat, and frequent travelers: a PLS impact-asymmetry analysis. *Tour Manag*. 2017;60:107–29.
49. Sharma A, Park S, Nicolau JL. Testing loss aversion and diminishing sensitivity in review sentiment. *Tour Manag*. 2020;77:104020.
50. Matzler K, Sauerwein E. The factor structure of customer satisfaction. *Int J Serv Ind Manag*. 2002;13(4):314–32.
51. Matzler K, Bailom F, Hinterhuber HH, Renzl B, Pichler J. The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance–performance analysis. *Ind Mark Manag*. 2004;33(4):271–7.
52. Mittal V, Ross WT, Baldasare PM. The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *J Mark*. 1998;62(1):33–47.
53. Schofield P, Coromina L, Camprubi R, Kim S. An analysis of first-time and repeat-visitor destination images through the prism of the three-factor theory of consumer satisfaction. *J Destin Mark Manag*. 2020;17:100463.
54. Shanmugam K, Jeganathan K, Mohamed Basheer MS, Mohamed MA, Firthows, Jayakody A. Impact of business intelligence on business performance of food delivery platforms in Sri Lanka. *Glob J Manag Bus Res*. 2020;20(6):39–51.
55. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. *J Exp Soc Psychol*. 1986;22(5):453–74.
56. Abbasi GA, Kumaravelu J, Goh YN, Dara Singh KS. Understanding the intention to revisit a destination by expanding the theory of planned behaviour (TPB). *Span J Mark - ESIC*. 2021;25(2):282–311.
57. Meng B, Cui M. The role of co-creation experience in forming tourists' revisit intention to home-based accommodation: extending the theory of planned behavior. *Tour Manag Perspect*. 2020;33:100581.
58. Kim JJ, Han H. Saving the hotel industry: strategic response to the COVID-19 pandemic, hotel selection analysis, and customer retention. *Int J Hosp Manag*. 2022;102:103163.
59. Schuckert M, Liu X, Law R. A segmentation of online reviews by language groups: how english and non-english speakers rate hotels differently. *Int J Hosp Manag*. 2015;48:143–9.
60. Guo Y, Barnes SJ, Jia Q. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. *Tour Manag*. 2017;59:467–83.
61. Liu Y, Teichert T, Rossi M, Li H, Hu F. Big data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tour Manag*. 2017;59:554–63.
62. Pizam A, Shapoval V, Ellis T. Customer satisfaction and its measurement in hospitality enterprises: a revisit and update. *Int J Contemp Hosp Manag*. 2016;28(1):2–35.
63. Lee M, Cai Y, DeFranco A, Lee J. Exploring influential factors affecting guest satisfaction. *J Hosp Tour Technol*. 2020;11:137–53.
64. Qiu L, Pang J, Lim KH. Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity: the moderating role of review valence. *Decis Support Syst*. 2012;54(1):631–43.
65. Valdivia A, Luzon MV, Herrera F. Sentiment analysis in tripadvisor. *IEEE Intell Syst*. 2017;32(4):72–7.
66. Kordzadeh N. "An empirical examination of factors influencing the intention to use physician rating websites," *Proc. Annu. Hawaii Int. Conf. Syst. Sci*, vol. 2019-Janua, pp. 4346–4354, 2019.
67. Koh NS, Hu N, Clemons EK. Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electron Commer Res Appl*. 2010;9(5):374–85.
68. Tsai C-F, Chen K, Hu Y-H, Chen W-K. Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tour Manag*. 2020;80:104122.
69. Chen Z, Jiang L, Li C. Label augmented and weighted majority voting for crowdsourcing. *Inf Sci (Ny)*. 2022;606:397–409.
70. Dong Y, Jiang L, Li C. Improving data and model quality in crowdsourcing using co-training-based noise correction. *Inf Sci (Ny)*. 2022;583:174–88.
71. Park E, Kang J, Choi D, Han J. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. *Curr Issues Tour*. 2020;23(5):605–11.
72. Guerreiro J, Rita P. How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *J Hospitality Tourism Manage*. 2020;43:269–72.
73. Davani AM, Díaz M, Prabhakaran V. Dealing with disagreements: looking beyond the majority vote in subjective annotations. *Trans Assoc Comput Linguist*. 2022;10:92–110.
74. Jain PK, Srivastava G, Lin JC-W, Pamula R. Unscrambling customer recommendations: a novel LSTM ensemble approach in airline recommendation prediction using online reviews. *IEEE Trans Comput Soc Syst*. 2022;9(6):1777–84.
75. Sharma A, Shafiq MO. A comprehensive artificial intelligence based user intention assessment model from online reviews and social media. *Appl Artif Intell*. 2022;36(1):2014193.
76. Xia R, Xu F, Yu J, Qi Y, Cambria E. Polarity shift detection, elimination and ensemble: a three-stage model for document-level sentiment analysis. *Inf Process Manag*. 2016;52(1):36–45.
77. Naldi M, Petroni S. A Testset-Based method to analyse the negation-detection performance of lexicon-based sentiment analysis tools. *Computers*. 2023. <https://doi.org/10.3390/computers12010018>.
78. Laurence R, Horn. "A Natural History of Negation," Univ Chic Press, 1989.
79. Sykes D, et al. Comparison of rule-based and neural network models for negation detection in radiology reports. *Nat Lang Eng*. 2021;27(2):203–24.

80. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform.* 2001;34(5):301–10.
81. Singh KN, Devi SD, Devi HM, Mahanta AK. A novel approach for dimension reduction using word embedding: an enhanced text classification approach. *Int J Inf Manag Data Insights.* 2022;2(1):100061.
82. Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf AJ. "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," pp. 3–21, 2020.
83. Hussein DME-DM. A survey on sentiment analysis challenges. *J King Saud Univ - Eng Sci.* 2018;30(4):330–8.
84. Amato F, Coppolino L, Cozzolino G, Mazzeo G, Moscato F, Nardone R. Enhancing random forest classification with NLP in DAMEH: a system for DData management in eHealth domain. *Neurocomputing.* 2021;444:79–91.
85. Nikolenko SI, Koltcov S, Koltsova O. Topic modelling for qualitative studies. *J Inf Sci.* 2017;43(1):88–102.
86. Churchill R, Singh L. The evolution of topic modeling. *ACM Comput Surv.* 2022. <https://doi.org/10.1145/3507900>.
87. Roberts ME, et al. Structural topic models for Open-Ended Survey responses. *Am J Pol Sci.* 2014;58(4):1064–82.
88. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell.* 1997;97:1–2.
89. Gumus M, Kiran MS. "Crude oil price forecasting using XGBoost," in 2017 International Conference on Computer Science and Engineering (UBMK), Oct. 2017, pp. 1100–1103.
90. Midtjord AD, De Bin R, Huseby AB. A decision support system for safer airplane landings: Predicting runway conditions using XGBoost and explainable AI. *Cold Reg Sci Technol.* 2022;199:103556.
91. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining Black Box Models. *ACM Comput Surv.* 2019;51(5):1–42.
92. Ahani A, Nilashi M, Ibrahim O, Sanzogni L, Weaven S. Market segmentation and travel choice prediction in spa hotels through TripAdvisor's online reviews. *Int J Hosp Manag.* 2019;80:52–77.
93. Liu J. "515K Hotel Reviews Data in Europe," *Kaggle*. Kaggle, 2017. [Online]. Available: <https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe/version/1>.
94. V. Major, A. Surkis, and Y. Aphinyanaphongs, "Utility of General and Specific Word Embeddings for Classifying Translational Stages of Research," *AMIA ... Annu. Symp. Proceedings. AMIA Symp.*, vol. 2018, pp. 1405–1414, 2018
95. Chen S, Wang X, Zhang H, Wang J, Peng J. Customer purchase forecasting for online tourism: a data-driven method with multiplex behavior data. *Tour Manag.* 2021;87:104357.
96. Christodoulou E, Gregoriades A, Pampaka M, Herodotou H. "Application of Classification and Word Embedding Techniques to Evaluate Tourists' Hotel-revisit Intention," in *Proceedings of the 23rd International Conference on Enterprise Information Systems*, 2021, vol. 1, no. Iceis, pp. 216–223.
97. Sharma DK, Pamula R, Chauhan DS. Query expansion – hybrid framework using fuzzy logic and PRF. *Measurement.* 2022;198:111300.
98. Roberts ME, Stewart BM, Tingley D. Stm: an R package for structural topic models. *J Stat Softw.* 2019;91:1–40.
99. Radojevic T, Stanic N, Stanic N. Inside the rating scores: a multilevel analysis of the factors influencing customer satisfaction in the Hotel industry. *Cornell Hosp Q.* 2017;58(2):134–64.
100. Banerjee S, Chua AYK. In search of patterns among travellers' hotel ratings in TripAdvisor. *Tour Manag.* 2016;53:125–31.
101. Berry LL, Parasuraman A, Zeithaml VA. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *J Retail.* 1988;64(1):12–40.
102. Lubo-Robles D, Devegowda D, Jayaram V, Bedle H, Marfurt KJ, Pranter MJ. "Machine learning model interpretability using SHAP values: Application to a seismic facies classification task," in *SEG Technical Program Expanded Abstracts 2020*, Sep. 2020, pp. 1460–1464.
103. Alsayat A. Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca. *Neural Comput Appl.* 2022;35(6):4701–22.
104. Vargas-Calderón V, Moros Ochoa A, Castro Nieto GY, Camargo JE. Machine learning for assessing quality of service in the hospitality sector based on customer reviews. *Inf Technol Tour.* 2021;23(3):351–79.
105. Xu J, Wang X, Zhang J, Huang S, Lu X. Explaining customer satisfaction via hotel reviews A comparison between pre- and post-COVID-19 reviews. *J Hosp Tour Manag.* 2022;53:208–13.
106. Albayrak T, Caber M. Prioritisation of the hotel attributes according to their influence on satisfaction: a comparison of two techniques. *Tour Manag.* 2015;46:43–50.
107. Hollebeek LD, Andreassen TW. The S-D logic-informed 'hamburger' model of service innovation and its implications for engagement and value. *J Serv Mark.* 2018;32(1):1–7.
108. Saura JR, Ribeiro-Soriano D, Palacios-Marqués D. Setting B2B digital marketing in artificial intelligence-based CRMs: a review and directions for future research. *Ind Mark Manag.* 2021;98:161–78.
109. Christodoulou E, Gregoriades A, Pampaka M, Herodotou H. "Evaluating the Effect of Weather on Tourist Revisit Intention using Natural Language Processing and Classification Techniques," in 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, 2479–2484.
110. Wachter S, Mittelstadt B, Russell C. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. *SSRN Electron J.* 1–52, 2017.
111. Devlin J, Chang MW, Lee K, Toutanova K. "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, 1, Mlm, pp. 4171–4186, 2019.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.