ORIGINAL PAPER

Gauging Airbnb review sentiments and critical key-topics by small area estimation

Luca Frigau[1] · Giulia Contu[1] · Marco Ortu[1] · Andrea Carta[1]

Abstract

In literature, several researchers have discovered that the reviews written about Airbnb accommodation tend to be extremely positive than those published on other famous platforms, consequently, many negative experiences remain untracked. Leaving negative experiences underrepresented hampers hosts' ability to improve their services. To overcome this gap, we employ Small Area Estimation to quantify negative sentiment in Airbnb reviews and the relative critical topics that characterize them. Our methodology involves a two-step process: first, we employ sentiment analysis and topic modeling to identify negative sentiment and critical issues, followed by the application of a mixed effect random forest model to provide a granular analysis of Airbnb reviews in small sub-populations in the context of small area estimation. We focus on domains of the city of Rome defined by geographical areas and the presence of hosts and Superhosts. Our findings reveal nuanced sentiment variations and critical topic proportions that traditional methods often overlook.

Keywords Small area estimation · Tourism data · Airbnb · MERF · NLP

✉ Giulia Contu
  giulia.contu@unica.it

  Luca Frigau
  frigau@unica.it

  Marco Ortu
  marco.ortu@unica.it

  Andrea Carta
  andrea.carta88@unica.it

[1] Department of Economics and Business Sciences, University of Cagliari, Viale Fra Ignazio 17, 09123 Cagliari, Italy

🖄 Springer

1 Introduction

In time, online reviews have become one the most relevant tools to evaluate the performance and the services offered by tourism facilities. The reviews have been considered a crucial instrument to build trust, *to help determine how listings are ranked in response to user queries* (Zervas et al. 2021 p. 2). Moreover, they are intrinsic indicators of the quality of the service offered by the hosts (Lawani et al. 2019). Finally, they have a stronger effect on the average rating expressed by the guest at the end of the Airbnb experience, and on the occupancy rate (Qiu et al. 2018; Contu et al. 2023).

In literature, particular attention has been focused on the evaluation of reviews related to tourism accommodation and, particularly, on the reviews written concerning Airbnb accommodation.

Different researchers have investigated the role of the reviews on the Airbnb phenomenon. They have highlighted that the reviews written about Airbnb accommodation tend to be extremely positive than those published on other famous platforms [see for instance Bridges and Vásquez (2018), Santos et al. (2020), Bulchand-Gidumal and Melián-González (2020), Zervas et al. (2021)]. Moreover, others have argued that the Airbnb review system seems to encourage positive reviews (see, for instance, Guttentag (2019)). This positive rating is generated firstly by the positive experience lived by guests; and secondly by the minimization of the negative experience. The last is due to different factors such as sociocultural norms of politeness, established trust among host and guest, lack of anonymity, the removal of reviews that violate their guidelines directly by Airbnb (Bridges and Vásquez 2018).

Additionally, Bulchand-Gidumal and Melián-González (2020) have highlighted that Airbnb guests do not rate when their experience is not positive. They do not want to harm the host with whom they have had personal contact, and do not want a negative review to appear on their profile.

Consequently, many negative experiences remain untracked, preventing hosts from being able to understand which negative elements distinguish their accommodation and therefore, what to invest in to improve the experience of their guests. This study aims to investigate the negative elements that characterize different domains of the city of Rome. The analysis is carried out by estimating their sentiment and relative critical topics of reviews in several small domains in order to provide the key elements that can impact negatively the Airbnb experience for hosts. Normally, the number of negative reviews may be too small to generate accurate estimates from the data. To overcome this issue, the Small Area Estimation method is adopted.

We employed Natural Language Processing techniques such as Sentiment Analysis and Topic Modeling to first define the sentiment of the reviews and categorize them into positive and negative ones. Then, we examined negative reviews in order to extract the critical issues (topics). We focus the analysis on the Airbnb reviews published for the accommodations located in the city of Rome. We found that our approach reveals nuanced variations in sentiment and critical

topic proportions that are often missed by conventional methods. The findings have practical implications for both hosts and the Airbnb platform in enhancing customer experience. Five sections, besides the introduction, complete this study. The second Section is focused on the review of the literature related to the methodologies used in the analysis. The third is related to methodology, and the fourth is to the presentation of the data. The results and implications are explained in the fourth part. Finally, the last Section focuses on concluding remarks, limitations, and future developments.

2 Background

2.1 Small area estimation

Small area estimation (SAE) is a statistical methodology that addresses the challenges of making accurate and reliable inferences about specific sub-populations (age, sex, etc.) or geographic areas (province, municipality, etc.) when sample sizes are limited (Rao and Molina 2015). It is emerging as an essential instrument in a variety of fields, including economics, public health, environmental science, tourism policy, and market research, where granular, localized information is critical (Zhang et al. 2014; Giusti et al. 2017; Petrucci and Salvati 2006; Orford 2010; Rosselló et al. 2021).

The fundamental challenge that SAE aims to address is the inherent variability in survey data when combined with smaller geographic units or sub-populations. SAE method includes direct and indirect estimators (Hidiroglou 2007). A direct estimator is one that solely takes values of the target variable from the domain of interest's sample units. In this category, we find the Horvitz–Thompson HT estimator (Horvitz and Thompson 1952), which is the simplest direct estimator and can account for sources of unequal selection probabilities (Hidiroglou 2007). Other direct estimators, used if auxiliary information is available at the population level, are the Generalized Regression Estimator (GREG) (Deville and Särndal 1992) and its variants domain-specific GREG and the survey regression estimator (Battese et al. 1988).

In the case of indirect estimators, values of the variable of interest are taken from domains and/or time periods that are distinct from the domain and time period of interest. These kinds of models can be either based at the area or unit level (Battese et al. 1988). They are specified at the area level if they link small area means of interest to area-specific auxiliary variables, whilst unit-level models link unit values of interest to unit-specific auxiliary variables, in general, area-level estimators are computed if the unit-level data are not available (Rao and Molina 2015). One of the most widely used area-based levels was introduced by Fay III and Herriot (1979). This estimator is a linear combination of a direct and a synthetic estimator. On the other hand, the unit-level model originated with Battese et al. (1988), who introduced the nested error regression (NER) model. Regarding this category of models, the Best Linear Unbiased Predictor (BLUP) and its empirical version for small area parameters (EBLUP), differs because the variance components are obtained under the assumption of uncorrelated random area effects. In more recent

times, various machine learning models for SAE have been introduced. Bilton et al. (2017, 2020) estimate levels of deprivation, such as poverty, by using tree-based models (for classification and regression framework respectively) that take into account survey design and provide resampling strategies to compute standard errors. Moreover, Viljanen et al. (2022) compared the conventional structured additive regression model (STAR) with gradient boosting with the decision trees model (XGboost) to predict the well-being of the Netherlands population. The two models showed similar results, however, the researchers state that XGboost has the advantage that it does not need prior knowledge or model specification. Again on tree-based methods, Krennmair and Schmid (2022) proposes the use of mixed effects random forests (MERF) for estimating spatially disaggregated indicators for SAE. In contrast to tree-based models, Kontokosta et al. (2018) use neural networks and SAE techniques to estimate waste production at the building scale in New York City, while Parker and Holan (2023) propose a Bayesian unit-level random neural network model for informative sampling focused on multinomial data. When compared to standard SAE statistical models, machine learning methods have a lot of potential advantages, such as more precise predictions, quicker training times for large data sets, increased robustness to various data sets, and less effort and expertise from statisticians to design and implement (Viljanen et al. 2022).

2.2 Sentiment analysis

Sentiment analysis is a technique that employs natural language processing (NLP) to investigate and extract subjective texts comprising user opinions, preferences, and sentiments. This type of analysis can be carried out at various levels of granularity, such as a whole document or the individual words that characterize it as a whole. In general, sentiment analysis can be categorized into three sub-levels: (1) the document level, (2) the sentence level, and (3) the aspect level (Nanli et al. 2012). Nowadays, sentiment analysis is becoming increasingly important due to the growing availability of huge quantities of textual data produced by social media. As a result, researchers are still developing new methodologies and improving existing ones. The goal of sentiment analysis is to identify Subjectivity/Objectivity, Discrete Emotions, and Polarity in textual data. Because only subjective text contains sentiment information, subjectivity/objectivity detection concerns the primary identification of subjective versus objective text. Discrete Emotion Detection is a more refined analysis that seeks to identify emotions from human language, such as joy, love, and many more. Polarity detection attempts to provide a sentiment score to a given text that is either qualitative (positive/negative) or quantitative.

Sentiment analysis techniques are divided into two categories (Kaur et al. 2017; Liu 2012): (1) lexicon-based approaches and (2) Machine learning-based methods. Approaches based on a lexicon (Taboada et al. 2011) are further subdivided into Dictionary-based and Corpus-based approaches (Darwich et al. 2019). The former involves the use of a linguistically expert-created dictionary, while the latter involves the use of co-occurrence statistics in text corpora.

2.3 Topic modeling

Topic Modeling is an unsupervised learning method that recognizes underlying patterns and themes in a set of documents and facilitates their representation according to the frequency of the words that constitute them. Among the most conventional techniques for topic modeling, we can find the Latent Dirichlet allocation (LDA) (Blei et al. 2003), which is a hierarchical Bayesian model, and Non-Negative Matrix Factorization (NMF) (Févotte and Idier 2011), which use linear algebra for topic extraction. Both of these techniques describe a document as a bag-of-words and model every document as a collection of latent topics. Yet since these bag-of-word representations do not take into account the context of words in a sentence, bag-of-words input may not accurately represent documents. Furthermore, its effectiveness in analyzing social media data has been heavily criticized (Egger and Yu 2021; Sánchez-Franco and Rey-Moreno 2022). To overcome these issues, recently, new techniques such as Top2Vec (Angelov 2020) and BERTopic (Grootendorst 2022) have been introduced. Top2Vec is an algorithm that uses word embedding, which means that by vectorizing the text data, it is possible to locate semantically similar words (Egger and Yu 2022). On the other hand, BERTopic is a topic model that generates coherent topic representations in a three-step algorithm. It starts by creating document embedding with a pre-trained language model. Second, it lowers the dimensionality of the document embedding and generates semantically close clusters that represent each one as a single topic, and finally, it extracts the topic representation through a class-based variation of Term Frequency-Inverse Document Frequency (TF-IDF). In summary, Topic modeling is a valuable technique in NLP and text analysis, and its role is expected to grow significantly as the volume of digital data produced each day increases. This technique has been applied to various real-world problems, for instance, Ortu et al. (2022) has applied a topic modeling technique to infer university student satisfaction using their issues reports. Moreover, Moreno et al. (2023), applied a BERTopic algorithm to Airbnb customer reviews to identify service characteristics relevant to guest satisfaction.

3 Methodology

3.1 Notation of natural language process

Let us consider a collection of M documents $\mathcal{Q} = \{q_1, \ldots, q_M\}$, where each one is referred to a single review r posted by a user v indicated by the pair element $(r, v) \in \mathcal{O}$. Let us consider a collection of texts $\mathcal{T} = \{t_1, \ldots, t_T\}$ describing the observations included in $\mathcal{O} = \{1, \ldots, m\}$, such that they have a surjective relation $k : \mathcal{T} \rightarrow \mathcal{O}$. We consider k as known; consequently, it is possible to assign the texts to the corresponding observation. Specifically, each t is considered as a collection of m unordered reviews, that is $t = \{r_1, r_2, \ldots\}$, and a review j as a set of words,

that is $r_j = \{w_1, w_2, \dots\}$. The collection of all the reviews is represented by the set $\mathcal{R} = \bigcup_{i=1}^{T} \{\forall r \in t_i\}$.
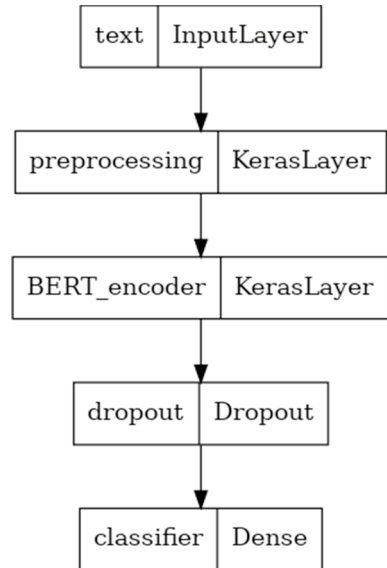
Our method consists of sentiment score assignment and topic extraction. First, the text documents are mapped to the set $\Omega = \{\mathcal{P}, \mathcal{N}\}$ such that $g : \mathcal{R} \to \Omega$, where g is a function that identifies the sentiment of a text, which can be either positive ($\mathcal{P}$) or negative ($\mathcal{N}$). The elements of $\mathcal{R}$ can be grouped according to their sentiment into a set of positive reviews $\mathcal{R}^+ = \{\forall r \in \mathcal{R} : g(r) = \mathcal{P}\}$ and into a set of negative ones $\mathcal{R}^- = \{\forall r \in \mathcal{R} : g(r) = \mathcal{N}\}$ such that $\mathcal{R}^+ \cup \mathcal{R}^- = \mathcal{R}$ and $\mathcal{R}^+ \cap \mathcal{R}^- = \emptyset$. The second step is the topic assignment. Let us define $\mathcal{W}$ as the collection of all words used in all reviews, that is the Bag-of-Words (BoW) of $\mathcal{R}$, and $\Theta = \{\theta_1, \dots, \theta_H\}$ as the set of all topics expressed in all reviews collected in $\mathcal{R}$. The BoW is transformed into a semantic vector representation using the Word-Embeddings (Goldberg and Levy 2014). The topic modeling method ϕ maps the Word-Embeddings representing the texts of reviews to topics such that $\phi : \mathcal{W} \to \Theta$ for each topic θ_h detected.

3.2 First step: sentiment analysis

We exploit BERT architecture (Devlin et al. 2018), one of the most effective models for NLP tasks (Rogers et al. 2021). Our foundation was a pre-trained large model designed explicitly for the English language. This choice ensured that the model had already learned complex language patterns, which significantly reduced the training time and potentially improved the accuracy when applied to specific tasks. To make the model suitable for the particular dataset of interest, we performed fine-tuning using a collection of documents with ratings (hotel reviews sourced from TripAdvisor or Booking.com, for instance). Online platforms, such as TripAdvisor and Booking.com, provide a unique rating system for their reviews, with scores rating s ranging from s_{min} to s_{max}. Let us consider as g function for sentiment analysis the following rule: reviews with a rating equal to or higher than a defined threshold s_τ are considered as "positive", while those with ratings less than s_τ are deemed "negative". This classification transforms a multi-level rating system into a dichotomous outcome, suitable for the fine-tuning of our model. To achieve a quantitative measure of sentiment, we integrated a softmax output layer in our neural network architecture as shown in Fig. 1. The initial layer accepts raw text as input, then the second layer is responsible for converting text data into a format that is usable by BERT, typically involving tokenization and converting tokens to embedding indices. The third layer is a pre-trained BERT model that processes the tokenized input and generates contextual embeddings for each token[1]. A dropout layer is then applied to the BERT encoder's output to prevent overfitting by randomly setting a fraction of the input units to 0 during training. Finally, a dense layer that serves as the classifier incorporating the softmax activation function to output probabilities over the target classes.

---

[1] In particular we used nlptown/bert-base-multilingual-uncased-sentiment.

Fig. 1 BERT fine-tuned deep neural network architecture

```
              ┌──────┬──────────────┐
              │ text │  InputLayer  │
              └──────┴──────────────┘
                        │
                        ▼
         ┌───────────────┬──────────────┐
         │ preprocessing │  KerasLayer  │
         └───────────────┴──────────────┘
                        │
                        ▼
         ┌───────────────┬──────────────┐
         │ BERT_encoder  │  KerasLayer  │
         └───────────────┴──────────────┘
                        │
                        ▼
             ┌─────────┬──────────┐
             │ dropout │ Dropout  │
             └─────────┴──────────┘
                        │
                        ▼
            ┌────────────┬─────────┐
            │ classifier │  Dense  │
            └────────────┴─────────┘
```

This layer transforms the model's raw output into a probability distribution over the two classes (negative and positive), implementing a surjective function $\psi : \mathcal{T} \rightarrow \mathbb{R}$. Thus, instead of merely providing a binary outcome, the model's final output is a continuous variable ranging between 0 and 1. An output of 0 indicates a negative sentiment, 1 denotes a positive sentiment, and values between the two give a probabilistic estimate of sentiment leaning.

3.3 First step: topic modeling

To delve deeper into the specific issues and critical aspects expressed in negative reviews, we employed topic modeling using the BERT architecture. A pre-trained large model specialized for the English language ensured that our model had inherent knowledge of the linguistic constructs. We first grouped the reviews based on sentiment. We employed the sentiment analysis model provided by the previous step, with a sentiment score between 0 and 1 for each review. Reviews that fell below a pre-defined threshold s_τ were categorized as negative and extremely negative. For topic modeling, we used the BERTopic model (Grootendorst 2022), which is particularly designed to leverage the strengths of the BERT architecture for extracting topics from textual data. This library aids in uncovering latent topics in data and highlighting negative aspects in reviews. BERTopic transforms each document into a high-dimensional vector that captures semantic information based on the context of words within the document using a pre-trained BERT model. Then, it applies a dimensionality reduction technique, such as Uniform Manifold Approximation and Projection (UMAP), to reduce the dimensionality of the embeddings while preserving their semantic relationships. Then it uses the Hierarchical Density-Based Spatial

Clustering of Applications with Noise (HDBSCAN) clustering algorithm to derive groups of documents that are interpreted as topics. Finally, it refines topics by using a class-based TF-IDF to extract representative words from each cluster. This helps in characterizing the topics identified through clustering with words distribution per topic.

To uncover patterns and recurrent themes that might be prevalent among dissatisfied customers, our methodology involves two phases. First, we used an unsupervised topic extraction to implement $\phi : \mathcal{W} \to \Theta$. Then, we manually inspect the obtained topics to derive a subset of seeded topics for a semi-supervised topic extraction $\phi_\lambda : \mathcal{W} \to \Theta$. The last phase ensures that the obtained topics are more coherent and interpretable to gauge the phenomenon under study.

3.4 Second step: small area estimation

Models for estimating area-level means from unit-level survey data are mostly regression-based linear mixed models (LMM), with random effects capturing the hierarchical structure of observations. Dealing with LMMs necessitates assumptions about the linearity and normalcy of random effects and error components, which are rarely supported by empirical evidence. According to Jiang and Rao (2020), the optimal results and prediction effectiveness of model-based SAE are inextricably linked to the validity of model assumptions. Estimates are potentially skewed and mean squared error (MSE) estimates are incorrect without theoretical and practical considerations for violated assumptions. Random forest is a data-driven prediction technique that combines flexible modeling properties without making explicit model assumptions (Breiman 2001). Thus, random forests help to strengthen models against model misspecification (Jiang and Rao 2020). It must adhere to basic property survey and inference theory, such as the management of hierarchically dependent data structures and uncertainty measures for derived indicators. For these reasons, we have chosen to use MERF by Krennmair and Schmid (2022) under unit-level population supplementary data.

Let us assume a finite population, denoted as U, consisting of N units distributed across D distinct domains $(U_1, U_2, \ldots, U_D)$. Each domain has its own set of units, with respective sizes $N_1, N_2, \ldots, N_D$. We are interested in estimating a continuous target variable, denoted as y_{ij}, for each individual observation j within each domain i, this target variable is available only for unit within the sample γ, which is drawn from U and consists of n units partitioned into sample sizes $n_1, n_2, \ldots, n_D$ for all D areas. Additionally, we have a vector of explanatory variables $\mathbf{x}_{ij}$ with p components for each unit in our sample. The relationship between $\mathbf{x}_{ij}$ and y_{ij} is modeled using a mixed effects regression framework:

$$y_{ij} = f(\mathbf{x}_{ij}) + u_i + e_{ij}. \tag{1}$$

Here, u_i follows a normal distribution with mean 0 and variance σ_u^2, and e_{ij} follows a normal distribution with mean 0 and variance σ_e^2. Defining f in Eq. (1) as a random forest produces a semi-parametric system that combines the benefits of random forests with the capacity to construct hierarchical survey data structures using random

effects. We, then, estimate area-level means with MERF (Hajjem et al. 2014), which enables the estimation of model-components f, u, σ_u^2, and σ_e^2 in the context of SAE. To fit f in Eq. (1), MERF algorithm proceeds in two main steps: (1) Estimating the forest function, assuming that the random effects term u_i is correct. (2) Estimating the random effects part, by assuming the Out-of-Bag predictions (OOB-predictions) from the forest to be correct.

Additionally, the variance components σ_e^2 and σ_u^2 are implicitly estimated by taking the expectation of maximum likelihood (ML) estimators given the data. We refer to Krennmair and Schmid (2022) for deeper methodological details. Finally, the final estimator for the area-level means for MERF ($\hat{\mu}_i^{\text{MERF}}$) is given by:

$$\hat{\mu}_i^{\text{MERF}} = \bar{\hat{f}}_i(\mathbf{x}_{ij}) + \hat{u}_i = \bar{\hat{f}}_i(\mathbf{x}_{ij}) + \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2/n_i} \left(\frac{1}{n_i} \sum_{j \in \gamma_i} \left(y_{ij} - \hat{f}(\mathbf{x}_{ij}) \right) \right)$$

$$\text{where } \bar{\hat{f}}_i(\mathbf{x}_{ij}) = \frac{1}{N_i} \sum_{j \in U_i} \hat{f}(\mathbf{x}_{ij})$$

(2)

This estimator avoids the problems associated with explicit model selection, which is prevalent in SAE. In reality, random forests, manage model selection inherently, including higher-order effects and non-linear interactions.

4 Case study

We carry out the analysis focusing on the reviews written for the Airbnb accommodations located in the city of Rome.

The reviews have been downloaded by InsideAirbnb.com, a website that provides data about Airbnb's impact on residential communities. Moreover, information related to Airbnb accommodations has been downloaded to comprehend which aspects can impact the sentiment and on negative topics.

On September 2023, we downloaded 923,375 reviews. After the sentiment evaluation, we found 2691 negative reviews relative to accommodation located in Rome. The time period of the reviews moves from the June 2012 to the June 2023. Few reviews are written in the first years. Only the 31.17 % of the reviews are written before the 2019. A reduced number has been written in 2020 during the COVID-19 pandemic. The 61.64% are written from the 2021 to the 2023.

We considered 116 geographical areas of Rome (Fig. 2) in the definition of the domains. In fact, the reviews are analyzed taking into account the different quarters of Rome, which has been divided into 116 non-administrative units: 22 districts, called in Italian rioni, located in the historic center of the city, 35 quarters surrounding the historic center of Rome, 6 suburbs, and 53 zones in the countryside of the city, as shown in Fig. 2. However, we have not found reviews for three areas, for this reason we carried out the analysis focusing on 113 areas.
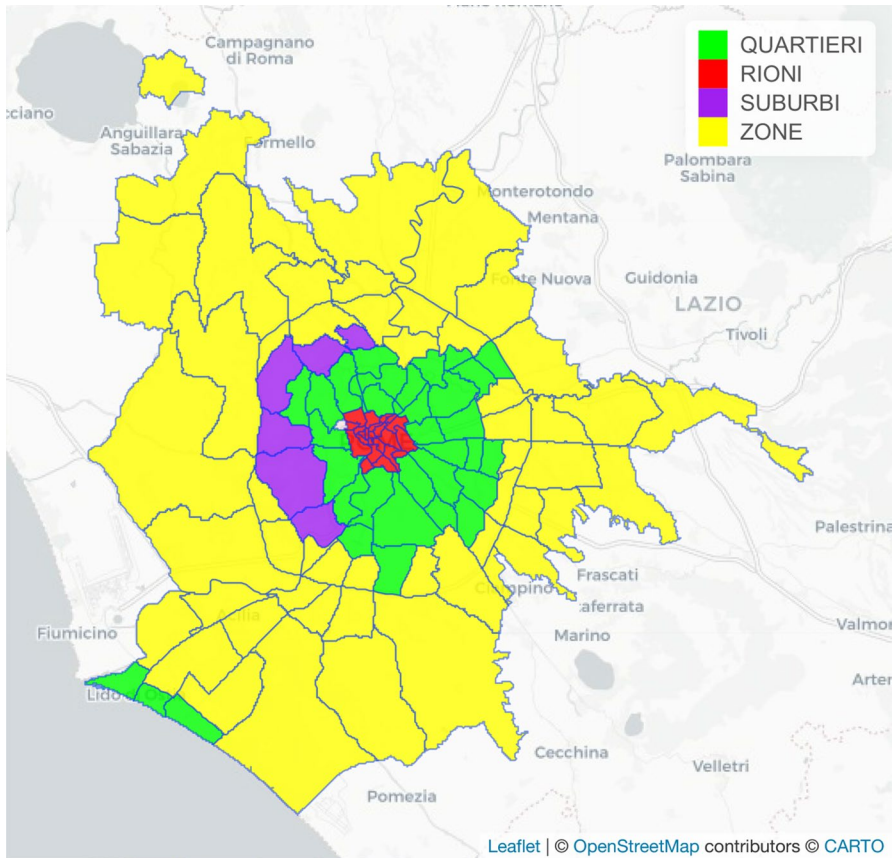
Fig. 2 Neighborhoods of Rome: in green (Quartieri), red (Rioni), purple (Suburbi), yellow (Zone)

Moreover, we decided to define the domains by combining two aspects: Superhosts and geographical areas. The Superhost status is a characterization of a host provided by Airbnb if in the previous 12 months, they have completed at least 10 stays, has a 4.8 or higher average overall rating, cancels less than 1% of the time, responds to 90% of new messages within 24 h.[2]

We decided to consider these two different Airbnb figures in the domains' definition because we hypothesize that Superhost manages the activity more professionally and, consequently, it can impact negative sentiment and topic.

Combining the 113 areas and the status of host and Superhost, we identify 226 subdomains. These subdomains are the object of the analysis. In literature, researchers have discovered that Airbnb accommodations are often close to the main tourist attractions [see for instance Gutiérrez et al. (2017), Celata and

---

[2] https://www.airbnb.com/e/superhost, accessed on April 29th, 2024.

Table 1 Fine-tuning of BERT model for sentiment analysis

Fine tuning dataset	Test dataset	Accuracy	Train size	Validation size	Test size
ACL-Imdb	ACL-Imdb	0.855	20,000	5000	25,000
ACL-Imdb	Tripadvisor	0.712	20,000	5000	4099
Tripadvisor	Tripadvisor	0.940	13,114	3278	4099

Table 2 Seeded words of the topics

Topic Label	Seeded words
Cleanness	Clean, dirty, cleaning, smell
Booking	Cancel, cancellation, reservation, booking, refund
Location	Place, location, street, centrally, center
Apartment	Room, shower, bathroom, kitchen, flat, door, sofa, mattress
Business ready	Internet, Wi-Fi, connection, conditioning, conditioner, connection
Host	Communication, host, Airbnb, contact, photo, comment
Experience	Noisy, noise, sleep, stay, experience, uncomfortable

Romano (2022)]; concentrated in specific areas of tourist cities [see for instance Dudás et al. (2017)]; and located in the most touristic areas of urban centers and in areas that are attractive and accessible by public transport (Quattrone et al. 2016). We hypothesize that the accommodations located in the center of the city are managed as tourism professional accommodation. For this reason, we suppose that differences in terms of sentiment and negative topics that influence this sentiment will be recorded among the different areas of Rome.

4.1 Sentiment and topic extraction

In order to estimate negative and critical reviews in the small domains considered we applied our methodology in three main steps: (1) Sentiment analysis, (2) Topic modeling, and (3) SAE and uncertainty estimation. In the sentiment analysis step, we first built a sentiment classifier using a pre-trained deep neural network with the BERT architecture (Devlin et al. 2018) shown in Fig. 1. In this phase, we exploited a pre-trained BERT model to extract the sentiment expressed in our Airbnb reviews dataset. We tested the pre-trained model on two datasets of reviews with ratings: Stanford ACL-Imdb (Tang et al. 2015) and Tripadvisor. com (Alam et al. 2016). We first used the Stanford ACL-Imdb datasets as it is a common choice for sentiment analysis on product or service reviews (Tang et al. 2015), obtaining an accuracy of 0.855. We tested the fine-tuned model on the Tripadvisor dataset, and we found an accuracy of 0.712. To improve this accuracy, we proceeded with a second fine-tuning on the Tripadvisor.com dataset with

Fig. 3 Refined topic words count

a final accuracy of 0.940 as summarized in Table 1. We set $s_\tau = 3$ to classify a review as positive ($s \geq 3$) or negative ($s < 3$), this value has been estimated empirically considering that the majority of reviews express a high rating as common in the touristic field (Bridges and Vásquez 2018; Santos et al. 2020; Bulchand-Gidumal and Melián-González 2020; Zervas et al. 2021).

Using our designed sentiment classifier we are able to cluster Airbnb reviews in two sets $\mathcal{R}^+ = \{\forall r \in \mathcal{R} : g(r) = \mathcal{P}\}$ and $\mathcal{R}^- = \{\forall r \in \mathcal{R} : g(r) = \mathcal{N}\}$ of positive and negative reviews. Since our focus is on critical issues expressed in the reviews, we considered the set of negative reviews $\mathcal{R}^-$ for topic extraction. For this purpose, we proceeded in two phases using the BERT pre-trained deep neural network BERTopic (Grootendorst 2022). In the first phase, we followed an unsupervised approach fitting the BERTopic model considering the negative reviews $\mathcal{R}^-$ set, obtaining 47 topics. We manually inspected these topics in order to reduce the number of topics to obtain a more interpretable set of topics. Table 2 shows the identified main topics with seeded words, we identified seven topics.

At this second stage, we refined the BERTopic providing the seven seeded topics as a base priors to identify topics. The final number of topics identified in this second stage was nine. Two out of these nine topics were removed after manual inspection. Figure 3 shows these nine topics with word bar plots. Each topic is represented by its top-5 most representative words. For example, Topic 8 represents critical issues about check-in/check-out, Topic 2 is about *Business Ready* problems with Wi-Fi and connection, Topic 7 is representative of issues with the location and the building, and so on. After a manual inspection of these topics, where we checked
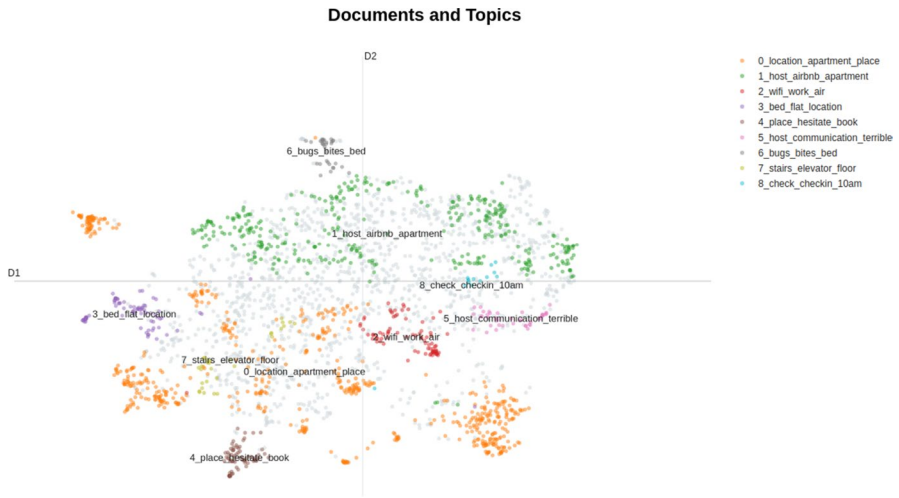
Fig. 4 Topic map after a two-dimension reduction

the reviews containing each topic, we chose to not consider Topic 1 and Topic 4. They refer to actual positive reviews, such as *"Don't hesitate to book"* for Topic 4, which were misclassified as negative reviews. Topic modeling is helpful in these cases due to its ability to identify latent structures in the corpus of reviews, allowing the refinement of the results of sentiment analysis.

Figure 4 shows a two-dimensional projection of documents, clustered by the most contributing topic, that preserves distances in the augmented document space, using t-distributed stochastic neighbor embedding (Van der Maaten and Hinton 2008). This figure shows well-separated clusters indicating that the topics identified are representative of documents that are closer in the augmented document space.

With these topics and the sentiment of reviews, we created the final dataset for the third step of our study where we applied MERF in order to estimate the model components f, u, σ_u^2, and σ_e^2 in the context of SAE. Our purpose is to estimate critical issues and sentiment means at the domain level.

4.2 Results

We fitted a MERF model in order to estimate the model-components f, u, σ_u^2, and σ_e^2 for the sentiment and topic probabilities variables. Each observation is characterized by 12 independent variables: *availability_365*, *bedrooms*, *beds*, *host_acceptance_rate*, *host_response_rate*, *host_response_time*, *instant_bookable*, *minimum_nights*, *price*, *property_type*, *room_type* and *verified* summarized in Table 3. In order to study the uncertainty, we compared the CV of the means per Sentiment and topic probabilities computed by MERF and by the Empirical Best Predictor (EBP) with a box-cox transformation (Molina and Rao 2010) from the R-package *emdi* (Kreutzmann et al. 2018). Furthermore, since the study of uncertainty plays an important role in SAE models, and CV measures the relative variability of the estimate of the
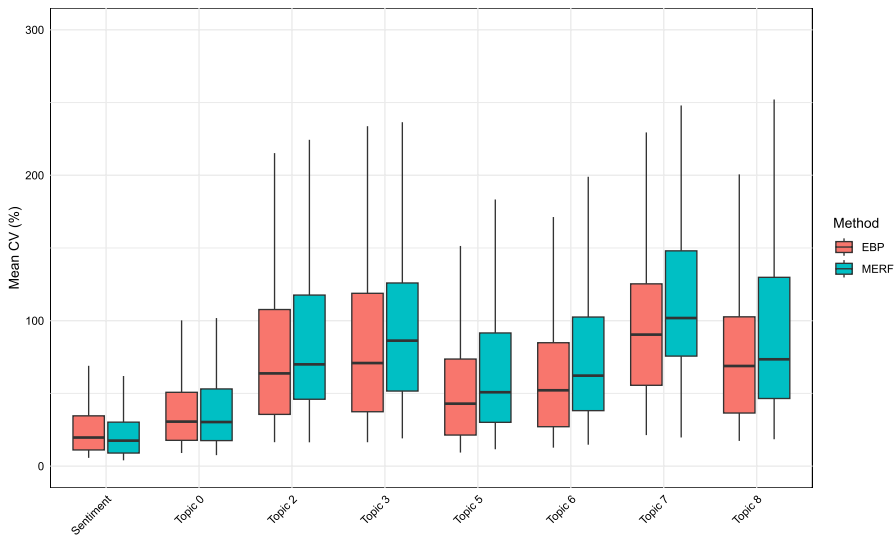
Fig. 5 Distributions of the coefficients of variation of the mean per Sentiment and probability to belong to the topics. In red the distributions computed by EBP, in green by MERF

Table 3 Description of the variables used in the analysis

Variable	Description
availability_365	Total number of listing calendar days, which were classified as available
bedrooms	Number of bedrooms in a vacation rental
listing	Three different listing types: entire accommodation, private room, and shared room
beds	Number of beds in the accommodation
host_acceptance_rate	The rate at which a host accepts booking requests
host_response_rate	Percentage of times a host responds to potential guests within 24 h
host_response_time	The amount of time that the guests must wait on average to obtain an answer to their questions
instant_bookable	Service that allows guests to book accommodation without explicit host approval
minimum_nights	Minimum night stay to book required by host
price	Available or booked nightly rate
property_type	Different types of accommodation on Airbnb
room_type	Private or shared room
verified	If the host has completed the ID verification process

mean for the specific domain compared to the true mean of the variable of interest, we decided to use this metric to compare the uncertainty reduction of the various models. For the MSE-bootstrap procedure, we use 100 repetitions for both EBP and MERF. About the convergence of MERF we set the error tolerance of MERF algorithm's convergence to 0.0001. If convergence was not achieved, the maximum

Table 4 Variable importance in relative terms as resulted by random forest models

	Sentiment	Topic 0	Topic 2	Topic 3	Topic 5	Topic 6	Topic 7	Topic 8
availability_365	0.255	0.262	0.254	0.289	0.255	0.285	0.326	0.307
bedrooms	0.048	0.042	0.040	0.050	0.045	0.034	0.034	0.018
beds	0.079	0.080	0.094	0.064	0.070	0.070	0.058	0.106
host_acceptance_rate	0.085	0.084	0.110	0.079	0.102	0.097	0.075	0.035
host_response_rate	0.077	0.066	0.049	0.049	0.067	0.069	0.055	0.074
host_response_time	0.018	0.016	0.020	0.012	0.014	0.028	0.007	0.010
instant_bookable	0.025	0.026	0.024	0.030	0.015	0.018	0.030	0.013
minimum_nights	0.069	0.078	0.057	0.066	0.075	0.039	0.047	0.031
price	0.249	0.254	0.271	0.267	0.268	0.294	0.321	0.313
property_type	0.076	0.068	0.065	0.077	0.075	0.061	0.042	0.066
room_type	0.014	0.014	0.015	0.014	0.013	0.005	0.004	0.026
verified	0.005	0.008	0.000	0.003	0.001	0.001	0.001	0.001

iteration number was set at 25. In the analysis, MERF converged in Sentiment, Topic 2, Topic 7, and Topic 8, respectively, after 4, 13, 3, and 7 iterations. Instead, the remaining models stopped at 25 iterations.

The uncertainty measured by MERF is comparable with that of EBP as it can be noted in Fig. 5. However being based on Random Forest MERF is intrinsically more flexible than EBP, which is a linear model, as well as being less sensitive to some of the critical assumptions of EBP. For this reason, the investigation is focused on the results of MERF. Moreover, we can also notice from Fig. 5 that the estimates for Sentiment are more precise than those of the probabilities to belong to the topics, in fact the values of the CV are much smaller for the variable Sentiment respect to the those of the variables that represent the probability of each review of belonging to a certain topic. However, this is consistent with expectations since identifying the topic of reviews is more complex than just measuring sentiment.

We examine the outcome of the variable importance from the random forest model in order to examine deeper into the model results. Table 4 illustrates this. This indicates that the price and the availability of the accommodations are the most significant factors that can affect the sentiment.

Moreover, the results underline lower values of sentiment for the domains characterized by the hosts (which we will call hosts for simplicity) than the domains characterized by the Superhosts (which we will call Superhosts for simplicity), as shown in Fig. 6.

The lowest value of sentiment for the hosts has been identified in the *suburbi Trionfale*, for the Superhost *Val Melania*. Both areas are far from the center of the city, and they are characterized by the presence of public houses. They are popular quarters. We suppose that these aspects can influence the sentiment expressed by the guests. On the contrary, the highest value for the hosts is recorded in *La pisana*, for the Superhost in *San Vittorino*. Both areas are residential and well connected to the city center.

Other interesting results have been obtained by the analysis of the probability that a specific topic can be included in the reviews. The results show differences in terms of probability to find a specific topics inside the domains. This means that this probability is affected by the domains and the location of the accommodation.

More in detail, Topic 0 is related to the apartment/location and the cleanness. Our results underline the relevance of this topic in negative reviews. It presents the highest probability with respect to the other topic. We noticed that the lowest value has been recorded in domain *San Basilio* for Superhost (0.14) and the highest one in the domain *Grottarossa* for host (0.38), as shown in Fig. 6. *Grottarossa* is a popular quarter, with problems related to safety and cleanness. These elements impact negatively on the reviews influencing the reviews of the Airbnb guests.

Topic 2 is related to the services called *Business Ready*. It identifies services that can support business travelers, such as Wi-Fi, a laptop-friendly workspace, iron, hangers, shampoo, hairdryer, and all the other essentials such as toilet paper and clean towels.

The results underline how this topic presents the lowest value in the domain *zona Val Melania* for Superhost and the highest value in the domain *Tor di Quinto*, as shown in Fig. 6. The lowest value in the first domain can be justified by the necessity of the Superhosts to offer better services in a popular area far from the center with respect to the other domain. The highest value in *Tor di Quinto* highlights how the worst services impact negative reviews and it suggests the necessity to improve the provided services.

Topic 3 is related to the sleeping quality and the location. This topic presents the lowest value in the domain *Torrenova* for Superhost (0.01) and the highest value in the domain *Tor Cervara* (0.15) for the host, as shown in Fig. 6. Both domains are far from the center of the city.

Topic 5 is related to the hosts and communication. When it is difficult to create positive communication and relationships between guests and hosts, the probability of finding this topic in negative reviews is elevated. This topic presents the lowest value in the domain *Fonte Ostiense* for the host (0.01) and the highest value in the domain *Tor San Giovanni* (0.18) for the host, as shown in Fig. 7.

Topic 6 can be called the *topic of Animals*. Finding bugs in the apartment can have a negative impact on the Airbnb experience. It negatively influences the reviews and the sentiment. The probability presents low values and a low range of variation. The lowest value has been recorded for the domain *San Basilio* for Superhost (0.01) and the highest value in the domain *Cecchignola* (0.09) for host, as shown in Fig. 7. Also, in this case, the probability of finding this topic is highest for the hosts.
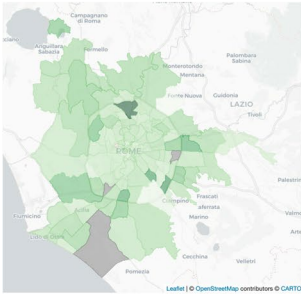
Topic 7 is related to the presence of an elevator and the problem of bringing luggage inside the apartment. Also, in this case, the probability presents low values and a low range of variation. More in detail, the lowest value has been recorded for
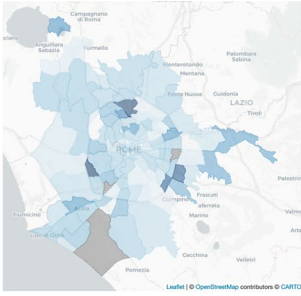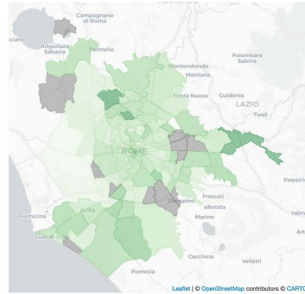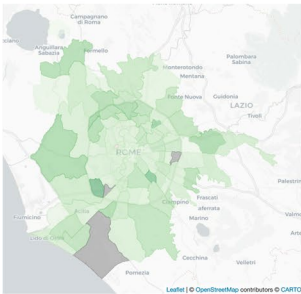
(a) *Sentiment - host*
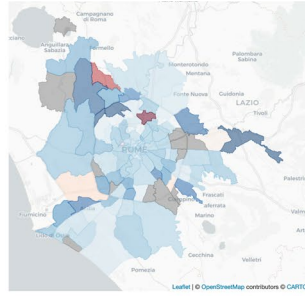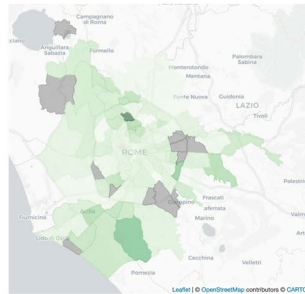
(b) *Sentiment - Superhost*
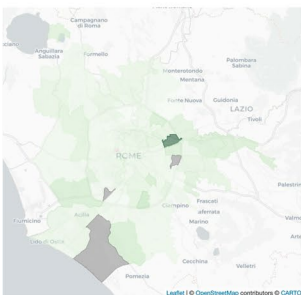
(c) *Topic 0 - host*
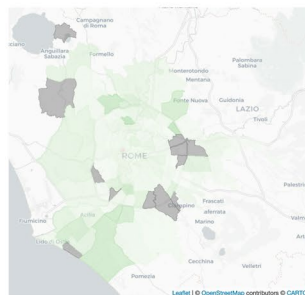
(d) *Topic 0 - Superhost*

(e) *Topic 2 - host*

(f) *Topic 2 - Superhost*
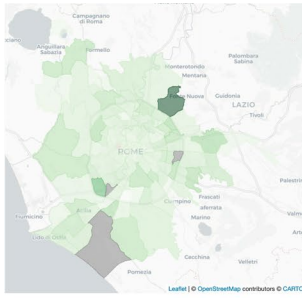
(g) *Topic 3 - host*

(h) *Topic 3 - Superhost*

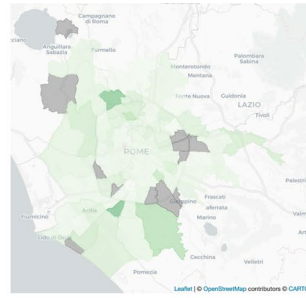(a) *Topic 5 - host*   (b) *Topic 5 - Superhost*
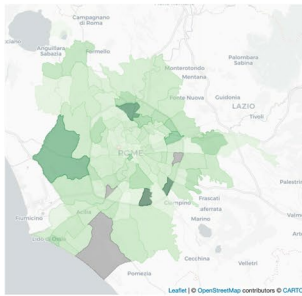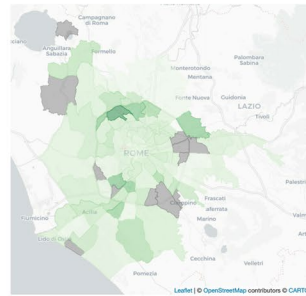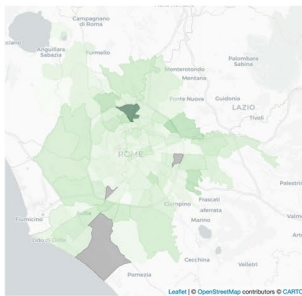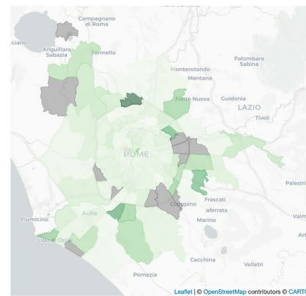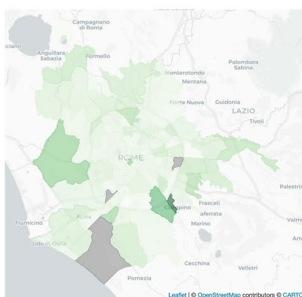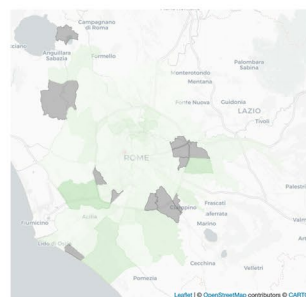
(c) *Topic 6 - host*   (d) *Topic 6 - Superhost*

(e) *Topic 7 - host*   (f) *Topic 7 - Superhost*

(g) *Topic 8 - host*   (h) *Topic 8 - Superhost*

Fig. 7 Distribution of the probability to belong to the Topics 5, 6, 7, 8 in the domains. Values range from light green (low) to dark green (high)
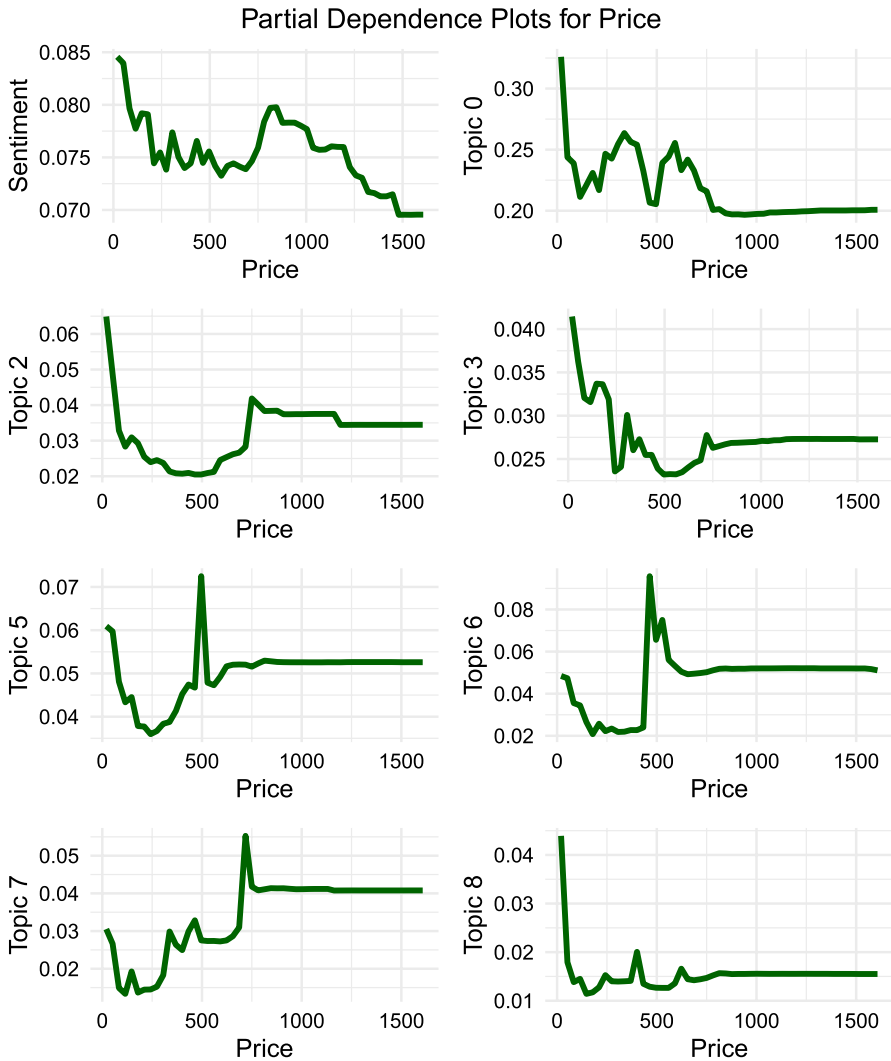
Fig. 8 Partial dependence plot for the variable *Price* for the MERF model

the domain *Fonte Ostience* for the host (0.001) and the highest value in the domain *Grottarossa* (0.07) for host, as shown in Fig. 7.

We know that the presence of an elevator is independent of the hosts, but it is more related to the building. However, we can suppose that Superhost has attempted to find supportive solutions.

Finally, Topic 8 is related to the check-in and the time of check-in.

The probability of Topic 8 presents low values and a low range of variation. More in detail, the lowest value has been recorded for the domain *Fonte Ostience* for the
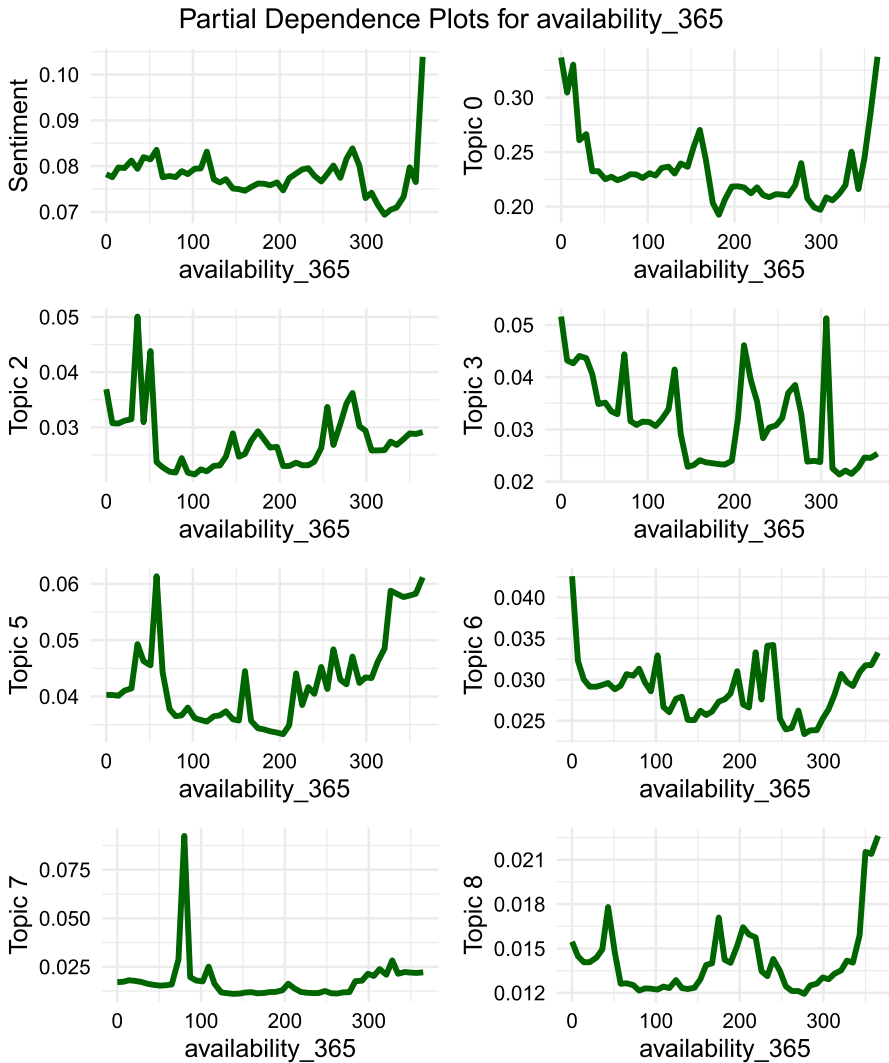
Fig. 9 Partial dependence plot for the variable *availability_365* for the MERF model

host (0.007) and the highest value in domain *Aeroporto di Ciampino* (0.08) for the host, as shown in Fig. 7.

Additionally, the results of the random forest applied on the topic show that the most important variables are the same record for the sentiment: their price and availability, as shown in Table 4. This means that the variable price influences more than others in the evaluation of the services.

To better understand the relationship between these two variables and the predicted outcome of the MERF, we computed their partial dependence plots (PDP). The PDP for *Price* and *availability_365* are shown in Figs. 8 and 9 respectively.

Figure 8 shows a not linear relationship between Price and Sentiment and each of the topics. More in detail, we can observe how an increase in the price generates a decrease in the estimated sentiment. Elevated prices are related to higher expectations. Consequently, it is possible to suppose that guests are more critical for accommodation with higher prices.

Moreover, it is interesting to underline how topics related to the location (Topic 0), the services (Topic 2), sleep quality (Topic 3), and check-in and check-out (Topic 8) reduce their probability of being discussed when the price increases. On the contrary, the aspects related to communication (Topic 5), the presence of insects (Topic 6), and the elevator (Topic 7) present a major probability to be discussed when the price increases. The results suggest that same elements as the possibility to communicate with the host are crucial for high elevate price.

For *availability_365* (Fig. 9), we still see a non-linear behavior for all the predicted outcomes. However, in contrast to the PDP of the variable price, also the predicted probability of belonging to any topic is influenced by the variable.

To sum up, we can state that firstly hosts present a more negative sentiment with respect to Superhost. Secondly, some topics are more negative in the reviews related to the host w.r.t. to Superhost. Thirdly, the highest probability is recorded for the topics related to the provided services, and the communication between host and guest. Finally, some domains exhibit more negative sentiment due to characteristics unrelated to them such as distance from the center.

4.3 Discussion

The results highlighted some relevant aspects that should be discussed in order to support the host and Superhost in the accommodation management. More in detail, it has been discovered that the price and the availability of the accommodation have the highest impact on the sentiment.

The relevance of the price has been previously underlined by Liang et al. (2017). They have argued that *an increase in the price of an accommodation offer will create more costs for guests to book it or post reviews of it. Thus, if the benefits are not enhanced correspondingly, guests will be less likely to book this property in order to make their choices rational. Especially for those guests with little prior knowledge of an accommodation offer, a higher price represents a greater risk.* (Liang et al. 2017 p. 457). In other words, a higher price creates higher guest expectations. Consequently, if the expectations are not reached, they generate a negative sentiment in the guest.

On the contrary, to our knowledge, it is the first time that the availability of accommodation has been considered as being able to impact the Airbnb sentiment. This result suggests that the guest prefers to have the possibility to book without limitation of days. Otherwise, it generates a negative sentiment and, consequently a negative review.

Similar results have been recorded in the identification of the topics. Also, in this case, the most important variables are the same record for the sentiment: their price and availability.

This means that the variable price influences more than others in the evaluation of the services. In fact, higher price generates higher expectations. If this equation is not respected, guest complains inside the review, taking into account specific aspects such as the cleanness and the provided services. Similarly, finding some limitations in the availability of accommodation can generate higher expectations, and, consequently, not finding specific kinds of services can generate more negative sentiment with respect to specific topics.

Moreover, the results suggest a more elevated negative sentiment for the hosts with respect to the Superhosts.

This statement is in line with the literature. For instance, Liang et al. (2017) have discovered that the accommodations managed by a Superhost are more likely to receive reviews than the normal hosts, and guests tend to post higher ratings for accommodations with a "Superhost" badge. Moreover, if we consider the domain in four macro areas, we can obtain an interesting and innovative result: the negative sentiment expressed by the guests is not influenced by the location of the accommodations. No differences have been recorded between hosts and guests operating in the different quarters of Rome as shown in Fig. 6. This suggests that the differences in terms of sentiment are more related to the provided services, and the characteristics of the accommodation. They are not related to the localization of the accommodations.

Additionally, we discovered that some topics are more relevant with respect to the others to be included in the reviews. For instance, the highest probability has been identified for the Topic 0. In fact, the apartment/location and the cleanness are considered as an element affecting the choice of the guests (Varma et al. 2016) and mentioned in negative experiences (von Hoffen et al. 2018).

Similar results have been recorded for Topic 2, highlighting how the judgments of the guests are related to the quality and quantity of the supplied services. These results have been previously highlighted in the literature. Guttentag et al. (2018) have stated that *Airbnb users are primarily attracted to the service by its practical advantages* (Guttentag et al. 2018 p. 354). Lalicic and Weismayer (2018) have argued that *hosts should be aware of the fact that guests request a specific standard, besides the interactive part of the hosting experience* (Lalicic and Weismayer 2018 p. 89).

Alike the relevance of Topic 3 is relevant. Previously in literature, it is been highlighted how bed comfort affects guests' satisfaction (von Hoffen et al. 2018), and how guests value positively comfortable beds (Tussyadiah and Zach 2017). Moreover, different researchers have evidence of how the guests prefer the location close to the city center [see for instance Quattrone et al. (2016)].

Also, in this case, the highest value has been recorded for the superhost with respect to the host. Moreover, the relevance of Topic 5 is related to the relationship between host and guest considered as the base of the Airbnb experience (Tussyadiah and Park 2018). It is crucial to have the possibility to speak with the host, also meeting him can generate a positive feeling and to influence positively

the perception of the Airbnb experience. The communication is required and has to be quick and immediate (Contu et al. 2023). For this reason, when it is difficult to create positive communication and relationships between guests and hosts, the probability of finding this topic in negative reviews is elevated.

Finally, the relevance of the aspects related to the check-in (Topic 8) has been previously evidenced in the literature. For instance, Ding et al. (2021) have stated the *checking in/out in negative reviews is often connected with Airbnb users' dissatisfaction that they are not able to check-in at a scheduled time and receive a flexible check-in or out services from the host* (Ding et al. 2021 p. 9).

Moreover, Alcañiz and Anguera-Torrell (2022) have stated that check-in practices affect guests' satisfaction in the context of vacation rental management companies. We can suppose that check-in is a topic recurrent in the reviews, and it can impact negatively for two reasons: firstly, the absence of contact in a crucial moment of the tourism experience; secondly, a non-adequate check-in time. The host should have a direct contact with the host and to be flexible with respect to the time.

5 Conclusions

The present study aimed to estimate the negative sentiment of the Airbnb accommodation located in Rome in specific domains defined by geographical areas and with respect to the presence of hosts and Superhosts. Since Airbnb reviews are, in most cases, positive, we tried to estimate the negative sentiment for areas where the negative reviews are few or totally absent. Moreover, we tried to comprehend which kinds of topics can influence the negative reviews and if differences exist between hosts and Superhosts, and with respect to the location of the accommodation. To reach this aim, we combined different methodologies: sentiment analysis, topic modeling, and SAE.

To our knowledge, this is the first study where SAE methodology is used to estimate the negative guests' sentiments and the topics that can impact the Airbnb experience. From a theoretical point of view, the use of MERF has greatly reduced the uncertainty of the negative sentiment as measured by the coefficient of variation.

From the managerial point of view, the analysis suggests which topics are relevant in negative reviews. Consequently, it guides the hosts in the identification of the elements that should be improved results in terms of sentiment. The topics with the higher probability are related to the services and the cleanliness. These two aspects are crucial in the Airbnb experience and should be improved to reach the best results in the reviews. Other aspects as the check-in and the presence of an elevator are less mentioned, but it should be important to consider them to offer a better service.

Moreover, the results confirm how the effort made to become a Superhost gives positive results in terms of sentiment. The host should work to obtain the Airbnb badge because it gives a major possibility to reach the best results in terms of guest sentiment. Finally, some domains should work to overcome objective limitations generated for instance by the distance to the center.

This paper presents some limitations. Firstly, we have applied the model considering only one city. It can be interesting to replicate the analysis on different cities in order to comprehend if the model can be useful in estimating the negative sentiment of Airbnb accommodation. Secondly, we focused on the Airbnb phenomenon, but it can be interesting to replicate the analysis for other tourism facilities or other kinds of services.

Declarations

References

Alam MH, Ryu WJ, Lee S (2016) Joint multi-grain topic sentiment: modeling semantic aspects for online reviews. Inf Sci 339:206–223

Alcañiz R, Anguera-Torrell O (2022) Vacation rental management firms: check-in practices and online ratings. Euro J Tour Res 30:1–13

Angelov D (2020) Top2vec: distributed representations of topics. arXiv:2008.09470

Battese GE, Harter RM, Fuller WA (1988) An error-components model for prediction of county crop areas using survey and satellite data. J Am Stat Assoc 83(401):28–36

Bilton P, Jones G, Ganesh S et al (2017) Classification trees for poverty mapping. Comput Stat Data Anal 115:53–66

Bilton P, Jones G, Ganesh S et al (2020) Regression trees for poverty mapping. Aust N Z J Stat 62(4):426–443

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3(Jan):993–1022

Breiman L (2001) Random forests. Mach learn 45:5–32

Bridges J, Vásquez C (2018) If nearly all Airbnb reviews are positive, does that make them meaningless? Curr Issue Tour 21(18):2057–2075

Bulchand-Gidumal J, Melián-González S (2020) Why are ratings so high in the sharing economy? evidence based on guest perspectives. Curr Issue Tour 23(10):1248–1260

Celata F, Romano A (2022) Overtourism and online short-term rental platforms in Italian cities. J Sustain Tour 30(5):1020–1039

Contu G, Frigau L, Ortu M (2023) Vglm proportional odds model to infer hosts' Airbnb performance. Qual Quant 57(5):4069–4094

Darwich M, Mohd SA, Omar N et al (2019) Corpus-based techniques for sentiment lexicon generation: a review. J Digit Inf Manag 17(5):296

Deville JC, Särndal CE (1992) Calibration estimators in survey sampling. J Am Stat Assoc 87(418):376–382

Devlin J, Chang MW, Lee K, et al (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805

Ding K, Choo WC, Ng KY et al (2021) Exploring sources of satisfaction and dissatisfaction in Airbnb accommodation using unsupervised and supervised topic modeling. Front Psychol 12(659):481

Dudás G, Boros L, Kovalcsik T et al (2017) The visualisation of the spatiality of Airbnb in Budapest using 3-band raster representation. Geograph Techn 12(1):23–30

Egger R, Yu J (2021) Identifying hidden semantic structures in Instagram data: a topic modelling comparison. Tour Rev 77(4):1234–1246

Egger R, Yu J (2022) A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Front Sociol 7(886):498

Fay RE III, Herriot RA (1979) Estimates of income for small places: an application of James-stein procedures to census data. J Am Stat Assoc 74(366a):269–277

Févotte C, Idier J (2011) Algorithms for nonnegative matrix factorization with the β-divergence. Neural Comput 23(9):2421–2456

Giusti C, Masserini L, Pratesi M (2017) Local comparisons of small area estimates of poverty: an application within the Tuscany region in Italy. Soc Indic Res 131:235–254

Goldberg Y, Levy O (2014) word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. arXiv:1402.3722

Grootendorst M (2022) Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv:2203.05794

Gutiérrez J, García-Palomares JC, Romanillos G et al (2017) The eruption of Airbnb in tourist cities: comparing spatial patterns of hotels and peer-to-peer accommodation in barcelona. Tour Manage 62:278–291

Guttentag D (2019) Progress on Airbnb: a literature review. J Hosp Tour Technol 10(4):814–844

Guttentag D, Smith S, Potwarka L et al (2018) Why tourists choose Airbnb: a motivation-based segmentation study. J Travel Res 57(3):342–359

Hajjem A, Bellavance F, Larocque D (2014) Mixed-effects random forest for clustered data. J Stat Comput Simul 84(6):1313–1328

Hidiroglou M (2007) Small-area estimation: theory and practice. In: Proceedings of the survey research methods section, pp 3445–3456

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47(260):663–685

Jiang J, Rao JS (2020) Robust small area estimation: an overview. Ann Rev Stat Appl 7:337–360

Kaur H, Mangat V et al (2017) A survey of sentiment analysis techniques. In: 2017 international conference on I-SMAC (IoT in Social. Mobile, Analytics and Cloud)(I-SMAC), IEEE, pp 921–925

Kontokosta CE, Hong B, Johnson NE et al (2018) Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities. Comput Environ Urban Syst 70:151–162

Krennmair P, Schmid T (2022) Flexible domain prediction using mixed effects random forests. J R Stat Soc Ser C Appl Stat 71(5):1865–1894

Kreutzmann A, Pannier S, Rojas-Perilla N, et al (2018) emdi: estimating and mapping disaggregated indicators. R package version 1(3)

Lalicic L, Weismayer C (2018) A model of tourists' loyalty: the case of Airbnb. J Hosp Tour Technol 9(1):80–93

Lawani A, Reed MR, Mark T et al (2019) Reviews and price on online platforms: evidence from sentiment analysis of Airbnb reviews in Boston. Reg Sci Urban Econ 75:22–34

Liang S, Schuckert M, Law R et al (2017) Be a "superhost": The importance of badge systems for peer-to-peer rental accommodations. Tour Manage 60:454–465

Liu B (2012) Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, vol 5. Springer, Berlin, pp 1–167

Molina I, Rao JN (2010) Small area estimation of poverty indicators. Can J Stat 38(3):369–385

Moreno MR, Sánchez-Franco MJ, Tienda MDISR (2023) Examining transaction-specific satisfaction and trust in Airbnb and hotels. An application of bertopic and zero-shot text classification. Tour Manag Stud 19(2):21–37

Nanli Z, Ping Z, Weiguo L, et al (2012) Sentiment analysis: A literature review. In: 2012 international symposium on management of technology (ISMOT), IEEE, pp 572–576

Orford S (2010) Towards a data-rich infrastructure for housing-market research: deriving floor-area estimates for individual properties from secondary data sources. Environ Plann B Plann Des 37(2):248–264

Ortu M, Frigau L, Contu G (2022) Topic based quality indexes assessment through sentiment. Comput Stat 39(1):289–311

Parker PA, Holan SH (2023) Computationally efficient Bayesian unit-level random neural network modelling of survey data under informative sampling for small area estimation. J R Stat Soc Ser A Stat Soc 186(4):722–737

Petrucci A, Salvati N (2006) Small area estimation for spatial correlation in watershed erosion assessment. J Agric Biol Environ Stat 11:169–182

Qiu W, Parigi P, Abrahao B (2018) More stars or more reviews? Differential effects of reputation on trust in the sharing economy. In: Proceedings of the CHI

Quattrone G, Proserpio D, Quercia D, et al (2016) Who benefits from the" sharing" economy of airbnb? In: Proceedings of the 25th international conference on world wide web, pp 1385–1394

Rao JN, Molina I (2015) Small area estimation. Wiley, London

Rogers A, Kovaleva O, Rumshisky A (2021) A primer in bertology: What we know about how bert works. Trans Assoc Comput Linguist 8:842–866

Rosselló J, Sansó A, Virbickaitė A (2021) How local tourism managers can benefit from national surveys: estimating tourism and restaurant expenditures for small market segments. Curr Issue Tour 24(24):3433–3449

Santos G, Mota VF, Benevenuto F et al (2020) Neutrality may matter: sentiment analysis in reviews of Airbnb, booking, and Couchsurfing in Brazil and USA. Soc Netw Anal Min 10:1–13

Sánchez-Franco MJ, Rey-Moreno M (2022) Do travelers' reviews depend on the destination? an analysis in coastal and urban peer-to-peer lodgings. Psychol Mark 39(2):441–459

Taboada M, Brooke J, Tofiloski M et al (2011) Lexicon-based methods for sentiment analysis. Comput Linguist 37(2):267–307

Tang D, Qin B, Liu T (2015) Learning semantic representations of users and products for document level sentiment classification. In: Proceedings of the 53rd annual meeting of the association for computational Linguistics and the 7th international joint conference on natural language processing, vol 1: long papers, pp 1014–1023

Tussyadiah IP, Park S (2018) When guests trust hosts for their words: host description and trust in sharing economy. Tour Manage 67:261–272

Tussyadiah IP, Zach F (2017) Identifying salient attributes of peer-to-peer accommodation experience. J Travel Tour Mark 34(5):636–652

Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(11):2579–2605

Varma A, Jukic N, Pestek A et al (2016) Airbnb: exciting innovation or passing fad? Tour Manag Perspect 20:228–237

Viljanen M, Meijerink L, Zwakhals L et al (2022) A machine learning approach to small area estimation: predicting the health, housing and well-being of the population of Netherlands. Int J Health Geogr 21(1):4

von Hoffen M, Hagge M, Betzing JH et al (2018) Leveraging social media to gain insights into service delivery: a study on Airbnb. IseB 16:247–269

Zervas G, Proserpio D, Byers JW (2021) A first look at online reputation on Airbnb, where every stay is above average. Mark Lett 32:1–16

Zhang X, Holt JB, Lu H et al (2014) Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. Am J Epidemiol 179(8):1025–1033