



**HAL**  
open science

# Sentiment Analysis for Tourism Insights: A Machine Learning Approach

Kenza Charfaoui, Stéphane Mussard

► **To cite this version:**

Kenza Charfaoui, Stéphane Mussard. Sentiment Analysis for Tourism Insights: A Machine Learning Approach. *Stats*, 2024, 7 (4), pp.1527 - 1539. 10.3390/stats7040090 . hal-04876406

**HAL Id: hal-04876406**

**<https://hal.science/hal-04876406v1>**

Submitted on 9 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

# Sentiment Analysis for Tourism Insights: A Machine Learning Approach

Kenza Charfaoui <sup>1,†</sup> and Stéphane Mussard <sup>1,2,\*</sup> 

<sup>1</sup> Faculty of Governance, Economics and Social Sciences, Mohammed VI Polytechnic University, Rabat 11100, Morocco; kenza.charfaoui@um6p.ma

<sup>2</sup> CHROME, University of Nîmes, Avenue du Dr. Georges Salan, 30000 Nîmes, France

\* Correspondence: stephane.mussard@unimes.fr

† These authors contributed equally to this work.

**Abstract:** This paper explores international tourism regarding Morocco's leading touristic city Marrakech, and, more precisely, its two prominent public spaces, Jemaa el-Fna and the Medina. Following a web-scraping process of English reviews on TripAdvisor, a machine learning technique is proposed to gather insights into prominent topics in the data, and their corresponding sentiment with a specific voting model. This process allows decision makers to direct their focus onto certain issues, such as safety concerns, animal conditions, health, or pricing issues. In addition, the voting method outperforms Vader, a widely used sentiment prediction tool. Furthermore, an LLM (Large Language Model) is proposed, the SieBERT-Marrakech. It is a SieBERT model fine-tuned on our data. The model outlines good performance metrics, showing even better results than GPT-4o, and it may be an interesting choice for tourism sentiment predictions in the context of Marrakech.

**Keywords:** Marrakech; large language models; machine learning; sentiment analysis; voting model

## 1. Introduction

### 1.1. Research Background

Marrakech, often referred to as the “red city”, is one of Morocco's most renowned destinations, with tourism being one of the cornerstones of the national economy, contributing significantly to the GDP and employment [1,2]. At the heart of this dynamic city are Jemaa el-Fna and the Medina, two cultural and historical landmarks. Jemaa el-Fna is a vibrant public square known for its street performances, food stalls, and artisan markets, while the Medina offers a labyrinth of narrow streets filled with centuries-old architecture and local crafts, all of which together embody the heritage of Morocco and contribute to the tourist experience.

Despite the popularity of these public spaces, their management presents unique challenges. The rapid influx of international tourists puts considerable pressure on local infrastructure, necessitates careful coordination to preserve cultural authenticity, and requires constant efforts to balance the needs of residents with the expectations of visitors. Effective responses to these concerns, therefore, require an in-depth understanding of tourist experiences. For this reason, online platforms such as TripAdvisor have become indispensable sources of information, offering first-hand accounts of visitors' interactions with these spaces. By analyzing the sentiments and themes expressed in these reviews, policymakers can identify priority areas for improvement and adapt their strategies to increase visitor satisfaction.

### 1.2. Research Topic: Exploring Tourist Sentiments on TripAdvisor

The rise in online customer opinions on specialized platforms and social media networks has led to an increased need for automated systems capable of organizing and classifying user reviews based on domain-specific aspects and sentiment polarities [3].



**Citation:** Charfaoui, K.; Mussard, S. Sentiment Analysis for Tourism Insights: A Machine Learning Approach. *Stats* **2024**, *7*, 1527–1539. <https://doi.org/10.3390/stats7040090>

Academic Editor: Wei Zhu

Received: 18 November 2024

Revised: 15 December 2024

Accepted: 18 December 2024

Published: 23 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

One platform that has emerged as an important repository for user-generated content in the field of tourism is TripAdvisor.

The literature highlights the practical importance of sentiment analysis on TripAdvisor by implementing a combination of topic modeling and sentiment analysis techniques. Focused on Marrakech, their study shows how aspect-based sentiment analysis enables an interesting exploration of each of the city's destinations' strengths and weaknesses [3]. The goal is then to provide tourism practitioners and specialists with actionable insights derived from user-generated content, positioning TripAdvisor as a useful source to understand the nuanced sentiments of tourists.

The discussion is extended to the economic aspect of tourism, specifically in the context of Morocco, through a study involving an exploratory data analysis of tourist sentiments toward Moroccan shopping places, using TripAdvisor reviews [4]. By employing data-mining techniques, the authors emphasize how a sentiment analysis built upon TripAdvisor's data can contribute not only to understanding subjective experiences but also to shaping economic dynamics within different touristic destinations.

The position of TripAdvisor in the tourism sector is further explored through the technical aspects of sentiment analysis on the platform. Research outlines the steps involved in sentiment classification on TripAdvisor, from polarity detection to aspect extraction and classification [5]. This highlights the technicalities of extracting sentiments from a vast pool of user-generated opinions, enhancing the platform's significance as a rich dataset available for textual analysis.

TripAdvisor's usability and vast potential are not simply attributed to its role as a review platform. It is also deeply linked to its co-creation ecosystem [6]. The willingness of users to actively contribute content and purpose the platform for trip planning emphasizes the mutually beneficial relationship between users and the platform. In this context, this serves not only as a repository of opinions but also as a dynamic space where users actively shape the content, showcasing its continuous value addition in the industry.

### *1.3. Tourism in Morocco*

The tourism landscape in Morocco has undergone significant evolution over time. Initially focused on attracting elite tourists with its oriental exoticism, Morocco's tourism sector witnessed a shift toward mass tourism in the 1970s, aligning with global trends. Despite initial skepticism toward the reorientation of tourism as a developmental tool, the state recognized its economic potential and began actively promoting the sector, welcoming international investments to stimulate its growth [7].

Regarding the economic aspects of tourism, the influx of tourists stimulates demand for locally produced goods and services, potentially leading to increased production and economic activity. In Morocco, the increase in international tourism revenues has been associated with positive effects on economic growth, demonstrating the potential of the sector as a driver of prosperity [1].

Recent statistics from the Moroccan Ministry of Tourism highlight the country's significant achievements in the tourism sector. In 2023, Morocco welcomed 14.5 million tourists, marking a significant increase of 34% compared to the year 2022 and 12% compared to 2019. Moroccan residents living abroad accounted for 51% of these arrivals, demonstrating a growth rate of 27% compared to 2022, while foreign tourists experienced a considerable surge of 41% over the same period (Ministère du Tourisme, de l'Artisanat et de l'Economie Sociale et Solidaire, <https://mtaess.gov.ma/fr/>, accessed on 24 April 2024).

In addition, Marrakech has a prominent position as the country's leading tourist destination. Rather than viewing the dominance of tourism in its economic profile negatively, it is considered a significant advantage for both the city itself and the national economy. Marrakech's status as an important tourist destination contributes significantly to its economic prosperity and further enhances Morocco's appeal on the global tourist stage [2].

As integral components of Morocco's tourism landscape, Marrakech's Jemaa el-Fna and the Medina are of profound significance as key attractions. Jemaa el-Fna, with its

lively atmosphere and rich cultural heritage, serves as a vibrant hub for non-locals and locals alike, offering an immersive experience into the country's traditions, cuisine, and entertainment. Similarly, the historic Medina, with its intricate alleys, ancient architecture, and vibrant souks, enchants visitors who come for its timeless charm and cultural opulence. These iconic landmarks not only showcase Morocco's cultural richness but also play pivotal roles in shaping the tourist narrative of Marrakech and the nation as a whole.

Furthermore, the TripAdvisor pages dedicated to Jemaa el-Fna and the Medina serve as repositories of invaluable insights, housing a plethora of reviews and experiences shared by visitors from around the world. These reviews offer nuanced perspectives and first-hand accounts of tourists' encounters with these iconic attractions, providing valuable information about visitor sentiments, preferences, and experiences.

#### 1.4. Main Results

The purpose of this paper is to explore the sentiments and topics in TripAdvisor reviews related to international tourism in Marrakech's prominent public spaces. This study employs a comprehensive methodology that leverages both machine learning (ML) techniques and artificial intelligence (AI) capabilities. To be precise, our research aims to determine the dominant themes and sentiments expressed by tourists who visit Jemaa el-Fna or the Medina.

1. We propose a voting method based on ML algorithms to analyze and predict the positive and negative sentiments prevalent in TripAdvisor reviews. This model extracts the most significant n-grams from tourism reviews in Marrakech. This technique is shown to be more accurate than other well-employed methods such as VADER (Valence Aware Dictionary and Sentiment Reasoner); see [8].

2. We provide a new Large Language Model (LLM), SieBERT-Marrakech, developed to improve sentiment analysis by predicting positive and negative sentiments in reviews.

3. An experimentation is conducted to compare each model with human annotators. SieBERT-Marrakech is shown to outperform the voting method, VADER, and GPT-4o, an advanced language model with enhanced efficiency and refined contextual understanding [9].

The paper is organized as follows. Section 2 provides a literature review on the relevance of studying online reviews in the Moroccan context. Section 3 presents our two models, the voting model and SieBERT-Marrakech, together with their results after being trained on 7958 positive sentiments and 1,437,309 negative sentiments. Section 4 provides a comparison of the models with human annotations. Finally, Section 5 proposes public policy orientations based on the results of voting and top features identification.

## 2. Literature Review

### 2.1. Marrakech's Tourist Attractions: Jemaa el-Fna and the Medina

Marrakech, with its famous Jemaa el-Fna and the Medina, has become an important point of interest, attracting tourists, investors, and new residents alike. Owning a riad or house in the Medina of Marrakech has transcended being a simple status symbol for the European elite and has become a trend embraced by a broader demographic. The inviting site, characterized by a favorable climate, accessibility, and panoramic views of the Atlas Mountains, adds to the charm of the city. Easy accessibility from major European cities further increases the desirability of Marrakech, as it is reachable within two or three hours by flight, not to mention how well the city is connected by road and train networks within Morocco [10].

However, while economically impactful, the advent of tourism entails challenges and potential negative consequences. Unregulated tourism can lead to the exploitation of resources and eclipse other economic sectors. Poorly managed tourism can also alter the essential balance for its growth, emphasizing the need for attentive public policies to mitigate the negative impacts [11].

Recognized as an influential tool for economic growth, tourism holds a significant place in Morocco's GDP. The utilization of spatio-temporal data, including social media data, is highlighted for tourism information needs. Understanding and incorporating emerging trends, especially digital ones, is therefore crucial for effective marketing and operational management in the sector [12].

However, the rise of westerners visiting Moroccan Medinas for vacation and residence, especially in Marrakech, introduces questions of cohabitation and identity [13]. These relationships become particularly evident when discussing identity, heritage value, and socio-urban practices of both new inhabitants and visitors. In addition, public spaces such as Jemaa el-Fna are not immune to tourism-driven transformations. However, these transformations often create a sort of gap between the projected and actual use of public spaces by residents, underlining the importance of understanding and respecting local practices [14]. The square is the subject of a dual discourse in the fields of tourist rhetoric and national rhetoric, but its patrimonialization aligns with both tourist and national narratives, highlighting its cultural significance [15].

Considering the above, the literature underscores the multidimensional importance of focusing on Marrakech, particularly spaces of interaction such as Jemaa el-Fna and the Medina. These spaces not only have cultural and historical significance but also represent key components of the Moroccan tourism industry, which makes them crucial for public policy considerations. Understanding the dynamics of tourism, cultural preservation, and interaction within these spaces provides a foundation for well-informed decision-making.

## 2.2. Online Reviews

To effectively highlight the relevance of online reviews as a data source, it is essential to begin with an understanding of TripAdvisor. TripAdvisor, established in 2000, has evolved into an extensive platform that aggregates travel-related information contributed by millions of users worldwide. Initially serving as a search engine that accessed travel data from various online sources, it became a hub for user-generated content, particularly reviews and ratings, which soon became the cornerstone of the platform [16]. Thanks to crowd-sourcing, TripAdvisor facilitates the diffusion of original and first-hand experiences shared by travelers, offering valuable insight into destinations, accommodations, and attractions.

For our study focusing on public spaces in Marrakech, specifically Jemaa el-Fna and the Medina, TripAdvisor reviews provide different narratives about visitors' interactions with the environment, such as the ambiance, cleanliness, and accessibility of public spaces. In addition, they offer information on social interactions within these spaces, including encounters with local vendors, cultural experiences, and safety concerns.

Empirical findings emphasize the importance of factors such as argument quality, source credibility, and perceived quantity of reviews to influence behavioral intentions [17]. Consumers rely on online reviews to gather informative and persuasive insights, helping them reduce uncertainty to make informed decisions.

In addition, studies have shown that evaluations of cognitive and sensory attributes in online reviews, such as hotel descriptions, exert a stronger impact on booking intentions and word-of-mouth recommendations compared to affective attributes [18]. Positive comments on sensory aspects, coupled with evaluations of cognitive elements, are particularly influential in shaping readers' intentions.

However, it is essential to point out potential concerns regarding the reliability of TripAdvisor reviews. Although studies indicate overall reliability, instances of questionable trustworthiness appear, highlighting the importance of critical evaluation when utilizing such data [19]. Consequently, TripAdvisor employs measures to mitigate issues such as fake reviews, including warnings against manipulation and penalties for offenders, which contributes to the credibility of the platform. Despite occasional challenges, the integrity and credibility of the platform system remain strong as evidenced by its effective moderation and the presence of genuine user-generated content [20].

### 3. Methodology

Two models of sentiment analysis are proposed. The first (Section 3.1) is based on machine learning algorithms (classifiers), which allow the best negative and positive words to be extracted with a voting scheme. The second (Section 3.2) is based on a new fine-tuned Large Language Model dedicated to treating sentiments of tourism in Marrakech.

#### 3.1. 'Voting' Method

Sentiment analysis is a particular application of supervised learning. It consists of assigning emotional labels to textual data, typically indicating polarity, i.e., whether the text yields positive or negative sentiment [21]. In supervised sentiment analysis, algorithms learn from labeled training data to infer a function that maps input text to sentiment labels, enabling the classification of unseen instances.

Supervised learning classifiers are employed to predict sentiment polarity based on the TF-IDF vectorization (unigrams and bigrams) derived from the reviews dataset [22,23]. This dataset is obtained through web-scraping techniques applied to the TripAdvisor pages specifically dedicated to each of the Medina and Jemaa el-Fna, capturing user-generated content related to these key attractions in Marrakech, along with the sentiment polarity corresponding to each review. To ensure robust classifier selection, we establish a criterion based on the F1 score, balancing precision and recall, with a threshold set at 70%. Classifiers above this threshold are kept, including the Ridge Classifier (RC), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Partial Least Squares Regression (PLSR), and Logistic Regression. We provide some brief comments about these classifiers below.

- Logistic Regression, based on maximum likelihood estimation for coefficient estimates, is commonly used for classification tasks.
- Ridge Classifier is based on the ridge regression incorporating a  $\ell_2$  regularization that minimizes overfitting, especially when dealing with limited dataset samples. The cost function includes a penalty term, with a higher penalization leading to more robust coefficients [24].
- Linear Discriminant Analysis (LDA) maximizes the between-class variance, aiming at reducing the dimensionality and improving the classification in a lower-dimensional subspace [25].
- Support Vector Machine (SVM) is a classifier that identifies hyperplanes to separate classes while avoiding overfitting [26].
- Partial Least Squares Regression (PLSR) is based on latent orthogonal latent components used to fit a linear regression model. PLSR involves determining the optimal number of latent components, typically through cross-validation [27].

These classifiers are fitted on the vectorized TF-IDF reviews (training data) with a binary target (positive/negative). The signs of the coefficient estimates are employed as polarity indicators, providing insight into the directionality of the relationship between features and predicted sentiment polarity. Positive coefficients therefore indicate a positive relationship, while negative coefficients suggest a negative one.

Subsequently, a dataframe is initialized to store the results of the voting process. Each column in this dataframe corresponds to a classifier used in the analysis (Table 1). For every feature in the dataset, the coefficients obtained from each classifier are examined. If a coefficient for a particular feature is greater than 0, signifying a positive influence, it is labeled as "Positive" for that classifier. Conversely, if the coefficient is negative or zero, indicating a lack of influence or a negative impact, the feature is labeled as "Negative".

A hard voting mechanism is used to determine the majority sentiment vote for each feature across all classifiers, without weighting particular classifiers (i.e., one vote for each classifier). The voting scheme is depicted in Figure 1 below.

As shown in Table 1, once each model assigns a label to each n-gram, the mode is computed, resulting in a final sentiment label ("Positive" or "Negative"). By aggregating the decisions of multiple classifiers, this ensemble learning approach provides a robust

and comprehensive assessment of the sentiment polarity associated with each feature in the dataset.

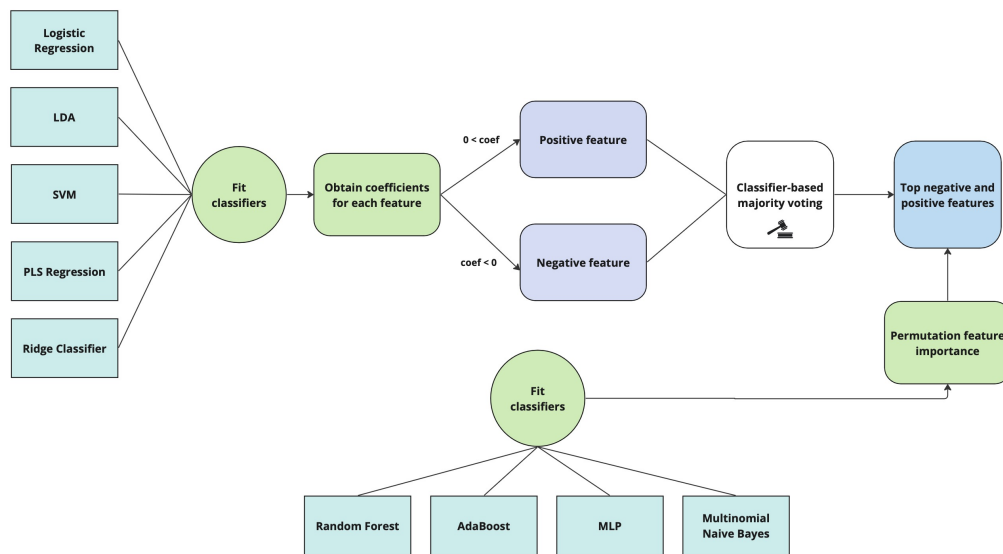


Figure 1. Voting and top features identification.

The n-gram polarity (positive/negative) is silent about the importance of a feature in a given model [28,29]. To identify the most influential features in the reviews, a permutation feature importance (PFI) methodology is applied. PFI provides the contribution of each feature to the overall model performance by randomly shuffling the values of one feature and observing the resulting accuracy fall.

Table 1. ‘Voting’ results (github link, <https://github.com/KenzaCH01/Voting-sentiment-analysis>, accessed on 29 May 2024).

n-Grams	Logistic Regression	Linear Discriminant Analysis	SVM	PLS Regression	Ridge Classifier	Final Decision
abuse	Negative	Negative	Negative	Negative	Negative	Negative
accept	Positive	Positive	Positive	Positive	Positive	Positive
acrobats	Positive	Positive	Positive	Positive	Positive	Positive
across	Negative	Negative	Negative	Negative	Negative	Negative
activity	Positive	Negative	Negative	Negative	Negative	Negative
...	...	...	...	...	...	...
worse	Negative	Negative	Negative	Negative	Negative	Negative
wrong	Negative	Negative	Negative	Negative	Negative	Negative
years	Positive	Positive	Positive	Positive	Positive	Positive
yes	Positive	Positive	Positive	Positive	Positive	Positive
young	Negative	Positive	Positive	Positive	Positive	Positive

716 rows × 6 columns.

The classifiers employed for the permutation feature importance are independent of those used for voting. These classifiers are the following.

- Multilayer Perceptron (MLP), a standard neural network architecture, stands out for its ability to deal with non-linear problems [30].
- AdaBoost is a meta-learning algorithm that builds on decision trees, combining weak learners to create a robust prediction rule [31,32].
- Random Forest, also an ensemble learning technique, aggregates decision trees using bootstrapping during the training process [33].
- The Naïve Bayes classifier relies on conditional probability and the assumption of feature independence to predict labels [34].

These classifiers were selected based on their ability to exceed a predefined performance threshold of 70%, ensuring that only models demonstrating robust predictive capabilities were retained for further analysis. Therefore, while PFI can be universally used with various classification and regression methods [35], the exclusion of classifiers such as K-Nearest Neighbors (KNN) and Decision Trees is necessary due to their relative inadequacy in achieving the predetermined performance threshold.

Subsequently, feature selection is performed to find the most important n-grams for each classifier independently using PFI. The intersection of these features across all classifiers is then computed to identify the most common n-grams among all models. The permutation feature importance has been compared with FESP; see [36]. FESP is a Shapley-based attribution method, which is faster than Shapley in extracting feature importance. However, on our dataset, FESP only displays negative features, in contrast to PFI, which displays both as shown in Table 2 below. Then, by identifying features shared among multiple classifiers, the permutation method brings out influential features across various ML approaches to obtain results based on the majority.

The ‘voting’ method, as previously described, is employed to determine the sentiment polarity (positive or negative) of each top feature (Table 2). This process (Figure 1) enables the extraction of the most salient features for sentiment analysis.

**Table 2.** Top features sentiment polarity.

Features	Sentiment
friendly	Positive
animal	Negative
night	Positive
harassed	Negative
overpriced	Negative
bill	Negative
fresh	Positive
money	Negative
charge	Negative
dirty	Negative
guide	Positive
lost	Positive
couldnt	Negative

### 3.2. Large Language Models: Fine-Tuning SieBERT

The Transformer architecture has emerged as a new paradigm in the field of natural language processing [37]. These models, unlike traditional recurrent neural networks (RNNs), rely on attention mechanisms computed over sequences to outline the global or local dependencies within input sequences [38]. To be precise, the self-attention mechanism captures the similarity between each couple of features in the same sentence, allowing the model to discern some relationships and patterns [39].

Transformer-based models rely on a pre-training phase on large datasets, followed by fine-tuning to specialize the model for specific tasks such as classification, summarization, etc. [39,40]. The pre-training phase, conducted in either a supervised or self-supervised manner, enables LLMs to better understand language structure and semantics, thereby improving their adaptability to numerous applications (see [39]).

The first strategy to train an LLM for positive and negative sentiments is to employ a pre-trained model such as RoBERTa [41] with 125 million parameters (Robustly Optimized BERT Pre-training Approach), which is not specialized in any specific task but just able to understand English. In order to capture the specialization of the model in sentiment analysis, a layer of  $768 \times 768$  is added (a preclassifier just before performing the binary classification). This allows new weights to be computed during the fine-tuning phase, resulting in a more specialized model dedicated to predicting positive and negative



sentiments. The dataset is composed of 7958 positive sentiments (label 0) and 1437 negative sentiments (label 1). The model is trained on 80% of the dataset (7516 reviews), while the remaining 20% is kept for testing. Over the training dataset, the model reaches a precision of 96.1%. Although the performance of this model is good, another strategy is chosen.

A specialized LLM for sentiment analysis, SieBERT, is developed and demonstrated to outperform RoBERTa across more than 15 datasets [42]. This is because SieBERT is trained to predict two labels (positive and negative sentiments) using RoBERTa Large, which consists of 355 million parameters. Our second strategy is to fine-tune SieBERT on our dataset. A layer of  $1024 \times 1024$  is added to SieBERT to obtain better specialization compared to our previous strategy. This new model fine-tuned on Marrakech reviews is called SieBERT-Marrakech. The training is carried out with batches of size 8, a learning rate of  $10^{-5}$  and 3 epochs (on one Nvidia RTX 8000).

The third strategy is to use a small model to measure the gap with large models. DistilBERT [43] has only 66 million parameters, making the fine-tuning phase of the model very fast while providing very good accuracy.

The results of the 3 models are reported below (Table 3: the results are rounded to 2 decimal points for RoBERTa and SieBERT-Marrakech for more precision).

**Table 3.** Metrics for 3 strategies on the testing set.

Metrics	RoBERTa	SieBERT-Marrakech	DistilBERT
Precision label 0	97.79%	98.04%	96%
Recall label 0	96.81%	96.69%	98%
F-measure label 0	97.30%	97.36%	97%
Precision label 1	82.65%	82.33%	83%
Recall label 1	87.41%	88.85%	76%
F-measure label 1	84.97%	85.47%	79%
F-measure (macro)	91.14%	<b>91.42 %</b>	88%

SieBERT-Marrakech displays better performance across multiple metrics compared to RoBERTa and DistilBERT. To access the SieBERT-Marrakech model interface, see the Huggingface hub ([https://huggingface.co/spaces/Steph974/Marrakech\\_sentiment\\_analysis](https://huggingface.co/spaces/Steph974/Marrakech_sentiment_analysis), accessed on 23 April 2024). The interface allows users to interact with the SieBERT-Marrakech model for sentiment analysis. Users can input text data, such as reviews or comments, into the interface, and the model predicts the sentiment of the input text, classifying it as positive or negative by providing probabilities for each class (Figure 2).



**Figure 2.** SieBERT-Marrakech usage demonstration ([https://huggingface.co/spaces/Steph974/Marrakech\\_sentiment\\_analysis](https://huggingface.co/spaces/Steph974/Marrakech_sentiment_analysis), accessed on 23 April 2024).

Although SieBERT-Marrakech is used for further experimentation (Section 4), it is important to note that all three models exhibit excellent performance. In particular, DistilBERT benefits from its compact size while still achieving good performance metrics.

#### 4. Experimental Results: Comparisons with VADER and GPT-4o

In this section, our aim is to compare the performance of the voting method with that of the SieBERT-Marrakech, VADER, and GPT-4o models for sentiment classification. Although the voting method is more inclined to select important features with their polarity, it can be employed as well as a classifier.

Since the voting method is designed for n-gram features rather than entire sentences such as reviews, a method is proposed to overcome this limitation. Positive and negative words in each review are computed, excluding noisy words, as those not present in the previous list of voted features (Table 1). If the number of positive words exceeds that of negative words, the review is classified as positive, and vice versa for negative classifications. In cases where the counts are equal, the review is labeled neutral. This process is applied on a balanced dataset to ensure an equal representation of both negative and positive labels in a sample of 100 reviews.

The SieBERT-Marrakech model is applied on the reviews to predict positive and negative labels. Then, Cohen's Kappa statistics is measured to obtain the inter-agreement between the voting method and SieBERT-Marrakech. Additionally, GPT-4o, OpenAI's latest version of the GPT-4 language model [9], is applied using the prompt 'I have comments of tourists who visited Marrakech. Give P for positive comments and N for negative. No neutral.' The Kappa score initially obtained when comparing voting and SieBERT-Marrakech is 0.5257, indicating medium inter-agreement. To account for the potential influence of the voting neutral ratings, these instances are excluded and the Kappa score is reassessed, resulting in a slightly improved score of 0.6429, while still suggesting moderate inter-agreement.

The Kappa inter-agreement is then computed between each method and two human annotators (see Table 4). The inter-agreement between the two human annotators displays a high Kappa score of 0.9599, indicating a strong agreement. When comparing each of the voting method and SieBERT-Marrakech with one of them, the Kappa scores are 0.5266 (0.6435 after removing neutral reviews) and 0.8985, respectively. This suggests that SieBERT-Marrakech outperforms the voting method, as it shows better agreement with humans. In addition, GPT-4o shows substantial agreement with both the human annotator (kappa score of 0.7182) and SieBERT-Marrakech (0.7768), while showing medium agreement with the voting (0.4719).

**Table 4.** Kappa statistics.

Voting vs. Human	SieBERT-Marrakech vs. Human	GPT-4o vs. Human	Human vs. Human
0.6435	0.8985	0.7182	0.9599

Our experimental study is extended to include a comparison with VADER (Valence Aware Dictionary and sEntiment Reasoner). VADER is a lexicon and rule-based sentiment analysis tool built upon a human-authored dictionary, in which words are annotated with their own sentiment, which is similar to the voting approach, and intensity scores [8]. The Kappa score between VADER and the voting method is quite low (0.3478). VADER also exhibits low inter-agreement with SieBERT-Marrakech and the human annotator.

To highlight each model's ability to predict positive and negative sentiments, the accuracy, precision, and recall scores are provided in Table 5, based on 90 randomly selected reviews. Based on these metrics, SieBERT-Marrakech is the better model for the prediction of positive (P) and negative (N) sentiments (see Table 6), outperforming both the GPT-4o and the voting method. Our voting method, although not competitive with SieBERT-Marrakech, outperforms VADER, which is largely used for market predictions and other tasks to predict sentiments.

The fine-tuning process, which yields a specialization of the model based on the linguistic characteristics of the data related to tourist experiences in Marrakech, allows SieBERT-Marrakech to provide more accurate sentiment predictions. Although simpler

than SieBERT-Marrakech, our voting method captures sentiment in a straightforward and interpretable manner, while still outperforming standard models such as VADER.

**Table 5.** Metrics of prediction \*.

Method	Precision	Recall	F-Measure
Voting	0.835	0.83	0.835
SieBERT-Marrakech	<b>0.96</b>	<b>0.97</b>	<b>0.965</b>
VADER	0.725	0.805	0.725
GPT-4o	0.895	0.89	0.89

\* Metrics on 90 reviews.

**Table 6.** Example of reviews and predictions \*.

Reviews	SieBERT-M	Voting	VADER	Human	GPT-4o
Too busy, too pushy, too dirty, too many people coming to you to sell stuff, in the evening I found it scary...gazillion wonderful restaurants	N	P	P	N	N
Personally I do not see the charm of this square... This is one of those places you visit just to be able to cross it off the list and say that you have seen it.	N	P	P	N	N
A fast paced city square with merchants of all kinds, unfortunately with much of the merchandise being the same tourist junk. Nice experience... Lots of food stands etc.	P	N	P	P	N
This probably the nr 1 must see place in Marrakech even though its authenticity has been gradually eroded... a lot of confusion everywhere.	P	P	N	N	P
You must visit the most crowed square in Africa! There is always someone making "noises"... Also don't "feed" the business of using monkeys or snakes.	P	N	P	P	N

\* P: positive, N: negative, SieBERT-M = SieBERT-Marrakech.

## 5. Discussion: Public Policy Implications

The top n-grams issued from our voting method (Table 2) provide valuable insights that can inform public policy in the tourism sector. For instance, positive sentiments associated with the feature "guide" suggests that the role of tour guides is highly regarded by tourists. This aligns with the efforts of the public authorities to support employees of tourist companies and guides through various initiatives, such as the provision of monthly allowances [44]. The positive sentiment toward guides indicates the potential success of these support programs and outlines the importance of investing in the professional development of guides to enhance the overall tourist experience. Positive features such as "friendly", "fresh", and "night" reflect aspects of the destination that appeal to tourists and contribute positively to their overall experience. Tourism strategies based on these positive features could include promoting cultural exchange programs, culinary experiences with fresh local produce, and improving communication about attractive nightlife options.

In contrast, n-grams related to shopping experiences, such as "souks" and "money", highlight the significance of commercial activities in tourism. The literature on bargaining practices in souks shows negative sentiments associated with "overpriced" and "charge" [45]. This suggests that addressing issues related to pricing transparency and fair-trade practices could contribute to a more positive shopping experience for tourists. Furthermore, the distinction between local and tourist interactions [46] underscores the need for policies that address pricing disparities. Addressing issues related to discrimina-

tory pricing practices, such as charging higher prices to tourists, can contribute to a more inclusive and equitable tourism environment.

In our sentiment analysis, the n-gram “animal” is associated with negative reviews, highlighting concerns about animal welfare and cruelty in tourism. Research shows the importance of recognizing the interests of animals in tourism and advocating for cooperative relationships with animal welfare organizations to address practices and ensure their well-being [47]. The negative impact of animal attractions, such as Barbary macaques, on the visitor experience also reflects broader concerns about animal welfare in tourism settings [48]. Addressing animal cruelty issues is imperative to promote ethical and sustainable tourism practices and to protect both animal welfare and tourist satisfaction.

The presence of the feature “dirty”, often associated with unclean and malodorous areas, as well as concerns regarding food safety and hygiene highlight the need for government intervention to ensure health and hygiene standards in tourism. The government plays an important role in the implementation of policies that promote a healthy and hygienic tourism sector, particularly in hotels and restaurants [49]. In addition, addressing issues such as improper waste management [50] is crucial for maintaining clean and attractive tourist areas. This promotes sustainable tourism practices and improves the reputation of a country as an attractive destination.

The negative sentiment associated with “harassed” requires proactive measures to ensure the safety and security of tourists, such as increasing police presence in tourist areas and providing comprehensive safety information to visitors, including emergency contact numbers, safety tips, and guidance on navigating unfamiliar surroundings [51].

## 6. Conclusions

This study analyzes the prevalent themes and sentiments relevant to international tourism in Marrakech’s two famous public spaces, Jemaa el-Fna and the Medina. Through the application of machine learning techniques, important features have been extracted that could potentially point out some underlying issues with these locations. Starting from safety problems to ethical issues involving animal cruelty, they all present individually relevant topics for further research. The impact of “tourist prices” on long-term sales dynamics could also be an important topic to tackle.

The voting method offers simplicity and interpretability, making it useful for platforms without traditional rating systems, such as YouTube comments or other text-based user-generated content. From a policy-making perspective, SieBERT-Marrakech also shows strong predictive performance, especially in handling nuanced sentiments in complex datasets such as TripAdvisor reviews. After fine-tuning, it achieves the best accuracy in sentiment classification, outperforming even state-of-the-art LLMs such as GPT-4o. The model could be useful for policymakers needing precise insights into tourist perceptions from large volumes of reviews.

Finally, the two proposed methods show better performance metrics than VADER. The use of SieBERT-Marrakech in a larger-scale analysis, which does not only focus on TripAdvisor, would be an interesting next step in understanding the tourism sector through the individual perspectives of tourists.

**Author Contributions:** All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Huseynli, N. Econometric Analysis of the Relationship Between Tourism Revenues, Inflation and Economic Growth: The Case of Morocco and South Africa. *Afr. J. Hosp. Tour. Leis.* **2022**, *11*, 135–146. [CrossRef]
- Rafik, K. The Tourism Sector and Territorial Development in Marrakech City-Morocco. *Int. J. Humanit. Educ. Res.* **2023**, *5*, 253–284. [CrossRef]
- Ali, T.; Marc, B.; Omar, B.; Soulaïmane, K.; Larbi, S. Exploring destination's negative e-reputation using aspect based sentiment analysis approach: Case of Marrakech destination on TripAdvisor. *Tour. Manag. Perspect.* **2021**, *40*, 100892. [CrossRef]
- Bouabdallaoui, I.; Guerouate, F.; Bouhaddour, S.; Saadi, C.; Sbihi, M. Advanced Exploratory Data Analysis for Moroccan Shopping Places in TripAdvisor. In *Advanced Research in Technologies, Information, Innovation and Sustainability*; Guarda, T., Portela, F., Augusto, M.F., Eds.; Springer Nature: Cham, Switzerland, 2022; pp. 257–271.
- Valdivia, A.; Luzon, M.V.; Herrera, F. Sentiment Analysis in TripAdvisor. *IEEE Intell. Syst.* **2017**, *32*, 72–77. [CrossRef]
- Yoo, K.H.; Sigala, M.; Gretzel, U. Exploring TripAdvisor. In *Open Tourism: Open Innovation, Crowdsourcing and Co-Creation Challenging the Tourism Industry*; Egger, R., Gula, I., Walcher, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 239–255.
- Almeida-García, F. Current issues of tourism in Morocco. In *Routledge Handbook of Tourism in Africa*; Routledge: London, UK, 2020.
- Hutto, C.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In Proceedings of the International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; Volume 8, p. 1.
- OpenAI. GPT-4o (Generative Pre-Trained Transformer 4 Optimized). 2024. Available online: <https://openai.com> (accessed on 14 December 2024).
- Escher, A.; Petermann, S.; Clos, B. Le Bradage de la Médina de Marrakech? In *Le Maroc à la Veille du Troisième Millénaire 7 Défis, Chances et Risques d'un Développement Durable*; Berriane, M., Kagermeier, A., Eds.; Actes du 6ème Colloque Maroc-Allemand de Paderborn 2000; Faculté des Lettres et des Science Humaines: Rabat, Morocco, 2001.
- Saddou, H. Tourisme à Marrakech; Impacts économiques, socioculturels et environnementaux éminents, *Espace Géogr. Soc. Marocaine* **2019**, *28/29*, 221–251.
- Steenbruggen, J. Tourism geography: Emerging trends and initiatives to support tourism in Morocco. *J. Tour. Hosp.* **2016**, *3*, 224–239.
- Kurzac-Souali, A.C. Rumeurs et cohabitation en médina de Marrakech: L'étranger où on ne l'attendait pas. *Hérodote* **2007**, *127*, 64–88. [CrossRef]
- Choplin, M.A.; Gatin, V. L'espace public comme vitrine de la ville marocaine: Conceptions et appropriations des places Jemaa El Fna à Marrakech, Boujoud à Fès et Al Mouahidine à Ouarzazate. *Noréis. Environ. Aménage. Soc.* **2010**, *214*, 23–40. [CrossRef]
- Gauthier, L. Jemaa El-Fna ou l'exotisme durable. *Géogr. Cult.* **2009**, *72*, 117–136. [CrossRef]
- Keates, N. Deconstructing TripAdvisor. *Wall Str. J.* **2007**, *1*, 1–6.
- Zhang, K.Z.K.; Zhao, S.J.; Cheung, C.M.K.; Lee, M.K.O. Examining the Influence of Online Reviews on Consumers' Decision-Making: A Heuristic-Systematic Model. *Decis. Support Syst.* **2014**, *67*, 78–89. [CrossRef]
- Roozen, I.; Raedts, M. The Effects of Online Customer Reviews and Managerial Responses on Travelers' Decision-Making Processes. *J. Hosp. Mark. Manag.* **2018**, *27*, 1–24. [CrossRef]
- Chua, A.Y.K.; Banerjee, S. Reliability of Reviews on the Internet: The Case of TripAdvisor. In Proceedings of the World Congress on Engineering & Computer Science, San Francisco, CA, USA, 23–25 October 2013.
- O'Connor, P. User-Generated Content and Travel: A Case Study on TripAdvisor.Com. In *Information and Communication Technologies in Tourism 2008*; O'Connor, P., Höpken, W., Gretzel, U., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 47–58.
- Stine, R.A. Sentiment Analysis. *Annu. Rev. Stat. Its Appl.* **2019**, *6*, 287–308. [CrossRef]
- Spärck Jones, K. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *J. Doc.* **1972**, *28*, 11–21. [CrossRef]
- Qaiser, S.; Ali, R. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *Int. J. Comput. Appl.* **2018**, *181*, 25–29. [CrossRef]
- Sayed, A.; Elgeldawi, E.; Zaki, A.; Galal, A. Sentiment Analysis for Arabic Reviews Using Machine Learning Classification Algorithms. In Proceedings of the International Conference, Aswan, Egypt, 8–9 February 2020; p. 63.
- Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassani, A.E. Linear Discriminant Analysis: A Detailed Tutorial. *AI Commun.* **2017**, *30*, 169–190. [CrossRef]
- Somvanshi, M.; Chavan, P.; Tambade, S.; Shinde, S.V. A Review of Machine Learning Techniques Using Decision Tree and Support Vector Machine. In Proceedings of the 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 12–13 August 2016; pp. 1–7.
- Krämer, N.; Sugiyama, M. The Degrees of Freedom of Partial Least Squares Regression. *J. Am. Stat. Assoc.* **2011**, *106*, 697–705. [CrossRef]
- Fumagalli, F.; Muschalik, M.; Hüllermeier, E.; Hammer, B. Incremental permutation feature importance (iPFI): Towards online explanations on data streams. *Mach. Learn.* **2023**, *112*, 4863–4903. [CrossRef]
- Molnar, C.; Freiesleben, T.; König, G.; Herbringer, J.; Reisinger, T.; Casalicchio, G.; Wright, M.N.; Bischl, B. Relating the Partial Dependence Plot and Permutation Feature Importance to the Data Generating Process. In *Explainable Artificial Intelligence*; Longo, L., Ed.; Springer Nature: Cham, Switzerland, 2023; Volume 1901, pp. 456–479.
- Nosratabadi, S.; Ardabili, S.; Lakner, Z.; Mako, C.; Mosavi, A. Prediction of Food Production Using Machine Learning Algorithms of Multilayer Perceptron and ANFIS. *Agriculture* **2021**, *11*, 5. [CrossRef]

31. Schapire, R.E. Explaining AdaBoost. In *Empirical Inference*; Schölkopf, B., Luo, Z., Vovk, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.
32. Kégl, B. *Introduction to AdaBoost*; Citeseer Publisher: Princeton, NJ, USA, 2009.
33. Reiss, I.; Baron, D.; Shahaf, S. Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets. *Astron. J.* **2019**, *157*, 16. [[CrossRef](#)]
34. Singh, G.; Kumar, B.; Gaur, L.; Tyagi, A. Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. In Proceedings of the 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, UK, 24–26 April 2019; pp. 593–596.
35. Kaneko, H. Cross-validated permutation feature importance considering correlation between features. *Anal. Sci. Adv.* **2022**, *3*, 278–287. [[CrossRef](#)] [[PubMed](#)]
36. Condevaux, C.; Harispe, S.; Mussard, S. Fair and Efficient Alternatives to Shapley-based Attribution Methods. In Proceedings of the ECMLPKDD 2022—The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Grenoble, France, 19–23 September 2022.
37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
38. Condevaux, C.; Harispe, S. LSG Attention: Extrapolation of pretrained Transformers to long sequences. In Proceedings of the PAKDD 2023—The 27th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Osaka, Japan, 25–28 May 2023.
39. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 1–41. [[CrossRef](#)]
40. Zhao, H.; Chen, H.; Yang, F.; Liu, N.; Deng, H.; Cai, H.; Wang, S.; Yin, D.; Du, M. Explainability for Large Language Models: A Survey. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–38. [[CrossRef](#)]
41. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.
42. Hartmann, J.; Heitmann, M.; Siebert, C.; Schamp, C. More than a Feeling: Accuracy and Application of Sentiment Analysis. *Int. J. Res. Mark.* **2023**, *40*, 75–87. [[CrossRef](#)]
43. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *arXiv* **2019**, arXiv:1910.01108.
44. OECD. *OECD Tourism Trends and Policies 2022*; OECD: Paris, France, 2022.
45. Kania, K.; Kałaska, M. Functional and spatial changes of souks in Morocco’s imperial cities in the context of tourism development. *Misc. Geogr.* **2019**, *23*, 92–98. [[CrossRef](#)]
46. Wagner, L.B. ‘Tourist Price’ and Diasporic Visitors: Negotiating the Value of Descent. *Valuat. Stud.* **2015**, *3*, 119–148. [[CrossRef](#)]
47. Fennell, D. Tourism and Animal Welfare. *Tour. Recreat. Res.* **2015**, *38*, 325–340. [[CrossRef](#)]
48. Stazaker, K.; Mackinnon, J. Visitor Perceptions of Captive, Endangered Barbary Macaques (*Macaca sylvanus*) Used as Photo Props in Jemaa El Fna Square, Marrakech, Morocco. *Anthrozoös* **2018**, *31*, 761–776. [[CrossRef](#)]
49. Rasethunsa, B. Health and Hygiene Strategies for Tourism Promotion: Guidelines for Africa. *J. Tour. Leis. Hosp.* **2022**, *4*, 158–164. [[CrossRef](#)]
50. Perkumienė, D.; Atalay, A.; Safaa, L.; Grigienė, J. Sustainable Waste Management for Clean and Safe Environments in the Recreation and Tourism Sector: A Case Study of Lithuania, Turkey and Morocco. *Recycling* **2023**, *8*, 4. [[CrossRef](#)]
51. Zou, Y.; Yu, Q. Sense of Safety Toward Tourism Destinations: A Social Constructivist Perspective. *J. Destin. Mark. Manag.* **2022**, *24*, 100708. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.