Use of ChatGPT and Tripadvisor data for discovering positive and negative aspects of the BioParque Vale Amazônia and the Carajás National Forest between 2014 and 2022

Uso de dados do ChatGPT e do Tripadvisor para descoberta de aspectos positivos e negativos do BioParque Vale Amazônia e da Floresta Nacional de Carajás entre 2014 e 2022

Uso de datos de ChatGPT y Tripadvisor para descubrir aspectos positivos y negativos del BioParque Vale Amazônia y del Bosque Nacional de Carajás entre 2014 y 2022

Alan Marcel Fernandes de Souza¹
Afonso Cesar Lelis Brandão²
Ítalo Flexa Di Paolo³
Thiago Nicolau Magalhães de Souza Conte⁴
Jéssica da Silva Miranda⁵
Wanderson Alexandre da Silva Quinto⁶

Abstract
The BioParque Vale Amazônia and the Carajás National Forest are tourist attractions that have their gateway in the city of Parauapebas, southeast of the State of Pará, approximately

¹ Doctor in Electrical Engineering at Universidade Federal do Pará (UFPA). Universidade Estadual do Pará (UEPA) - campus Castanhal. Castanhal, Pará, Brasil. E-mail: alan.souza@uepa.br
Orcid: https://orcid.org/0000-0002-1656-5714
² Doctoral student in Electrical Engineering at Universidade Presbiteriana Mackenzie (UPM). Universidade de São Paulo, São Paulo, São Paulo, Brasil. E-mail: afonso.brandao@prof.inteli.edu.br
Orcid: https://orcid.org/0009-0009-4814-9502
³ Doctoral student in Electrical Engineering at Universidade Federal do Pará (UFPA). Universidade Estadual do Pará (UEPA) – campus Ananindeua. Ananindeua, Pará, Brasil. E-mail: itflexa@uepa.br
Orcid: https://orcid.org/0000-0001-6968-6571
⁴ Doctor in Electrical Engineering at Universidade Federal do Pará (UFPA). Universidade Estadual do Pará (UEPA) - campus Castanhal. Castanhal, Pará, Brasil. E-mail: thiagonconte@uepa.br
Orcid: https://orcid.org/0000-0002-1288-366X
⁵ Doctoral Student in Electrical and Computing Engineering at Universidade Estadual de Campinas (UNICAMP). Universidade Estadual do Pará (UEPA) - campus Castanhal. Castanhal, Pará, Brasil.
E-mail: jessica.miranda@uepa.br Orcid: https://orcid.org/0000-0002-4446-9703
⁶ Doctor in Psychology at Universidade Federal do Pará (UFPA). Universidade Estadual do Pará (UEPA), campus Redenção. Redenção, Pará, Brasil. E-mail: w.quinto@uepa.br
Orcid: https://orcid.org/0000-0002-1573-1370
720 km from the capital, Belém. Many people visit these places daily, and a portion of them record reviews about their experiences on the internet. One of the platforms that collects such data is Tripadvisor, thus constituting it an important data source to understand the positive and negative aspects of these places, from the perspective of visitors. The aim of this study was to measure the degree of satisfaction and dissatisfaction of visitors to the BioParque and forest based on sentiment analysis, automatically conducted by ChatGPT, of 244 reviews extracted from Tripadvisor. Approximately 90% of the reviews for both places were classified as positive; about 4% as neutral, and around 6% as negative. These figures reflect how well-preserved the areas are. However, dissatisfaction regarding outdated information on the BioParque's website and the sanitation of some areas of the forest were reported. These results are valuable for supporting the management of these spaces and for the constant improvement of the visitor experience, that can monitor visitors’ perspectives over time, aiming to improve their experiences.

Keywords: Reviews. Sentiment Analysis. Artificial Intelligence. Data.

Resumo
O BioParque Vale Amazônia e a Floresta Nacional de Carajás são atrações turísticas que têm sua porta de entrada na cidade de Parauapebas, sudeste do Estado do Pará, a aproximadamente 720 km da capital, Belém. Muitas pessoas visitam esses lugares diariamente, e uma parte delas registra avaliações sobre suas experiências na internet. Uma das plataformas que recebe esses dados é o Tripadvisor, constituindo-se assim numa importante fonte de dados para compreender os aspectos positivos e negativos desses locais, na perspectiva dos visitantes. O objetivo deste estudo foi medir o grau de satisfação e insatisfação dos visitantes do bioparque e da floresta com base na análise de sentimento, realizada automaticamente pelo ChatGPT, de 244 avaliações extraídas do Tripadvisor. Aproximadamente 90% das avaliações de ambos os locais foram classificadas como positivas; cerca de 4% como neutro e cerca de 6% como negativo. Esses números refletem o quão bem preservadas as áreas estão. Porém, foi relatada insatisfação com informações desatualizadas no website do bioparque e com o saneamento de algumas áreas da floresta. Esses resultados são valiosos para apoiar a gestão desses espaços e para a melhoria constante da experiência do visitante, que pode monitorar as percepções dos visitantes ao longo do tempo, visando melhorar as suas experiências.

Use of ChatGPT and Tripadvisor data for discovering positive and negative aspects of the BioParque Vale Amazônia and the Carajás National Forest between 2014 and 2022

Resumen
El BioParque Vale Amazônia y el Bosque Nacional Carajás son atractivos turísticos que tienen su puerta de entrada en la ciudad de Parauapebas, al sureste del Estado de Pará, aproximadamente a 720 km de la capital, Belém. Mucha gente visita estos lugares a diario y una parte de ellos registra reseñas sobre sus experiencias en Internet. Una de las plataformas que recopila dichos datos es Tripadvisor, constituyéndose así en una importante fuente de datos para comprender los aspectos positivos y negativos de estos lugares, desde la perspectiva de los visitantes. El objetivo de este estudio fue medir el grado de satisfacción e insatisfacción de los visitantes del bioparque y del bosque a partir del análisis de sentimiento, realizado automáticamente por ChatGPT, de 244 reseñas extraídas de Tripadvisor. Aproximadamente el 90% de las reseñas de ambos lugares fueron calificadas como positivas; alrededor del 4% como neutral y alrededor del 6% como negativo. Estas cifras reflejan lo bien conservadas que están las zonas. Sin embargo, se reportó descontento por la información desactualizada en el sitio web del bioparque y el saneamiento de algunas áreas del bosque. Estos resultados son valiosos para apoyar la gestión de estos espacios y para la mejora constante de la experiencia del visitante, que puede monitorear las perspectivas de los visitantes a lo largo del tiempo, con el objetivo de mejorar sus experiencias.


Introduction

The city of Parauapebas-PA, located in the southeast of the State of Pará, approximately 720 km from the capital, had the BioParque Vale Amazônia (BVA) inaugurated in 1985, which covers an area of 30 hectares of native forest, promoting the free movement of various species such as birds, agoutis, and monkeys in its visitor-accessible areas. Tourists have the opportunity to engage in a unique immersive experience in the rich biodiversity of the Amazon rainforest. Additionally, the BVA supports initiatives for study, research, and environmental projects dedicated to the preservation of local biological diversity, showcasing and conserving species characteristic of the Amazonian fauna and flora (Inácio, 2023; Vale, 2023).

The Carajás National Forest (CNF), also with entry through the municipality of Parauapebas-PA, established in 1998 and covering about 400,000 hectares, also covers the...
Use of ChatGPT and Tripadvisor data for discovering positive and negative aspects of the BioParque Vale Amazônia and the Carajás National Forest between 2014 and 2022

municipalities of Parauapebas, Canaã dos Carajás, and Água Azul do Norte. Known for its rich environmental diversity, and plant and mineral extraction, the region gained attention in the 1980s due to its mineral abundance. Within the forest, scientific research and conservation, visitation, and protection activities also take place, with particular emphasis on studies of the harpy eagle and the hyacinth macaw, as well as birdwatching. The area hosts Brazil’s largest mining project, extracting iron, gold, manganese, granite, and copper on a daily basis. Attractions include a clear water waterfall, iron-rich caves, an industrial mining complex, birdwatching, forest trails, river canoeing, canga vegetation, a zoobotanical park, Serra Sul, rainwater ponds, and camping facilities (Inácio, 2023; GOV.BR, 2023).

Tripadvisor is an online travel platform that provides data and reviews by travelers about various tourist spots. Therefore, Tripadvisor can be considered a social network in the tourism sector, constituting an important source of data from various places that can be processed to discover knowledge about the services provided at these places (Sangkaew & Zhu, 2022; Dutra & Gosling, 2021).

ChatGPT is a language model developed by OpenAI based on the GPT (Generative Pre-trained Transformer) architecture. It functions as an AI (artificial intelligence) that engages in running text conversations and can understand a command (prompt) and generate human-like text in response. It has the ability to produce coherent and contextually relevant text based on pre-training on large amounts of diverse data.

The aim of this study is to measure the degree of satisfaction and dissatisfaction of visitors to the BVA and CNF based on sentiment analysis conducted by ChatGPT on reviews posted on Tripadvisor. To achieve this, the following research question must be answered: What are the positive and negative aspects of the BVA and CNF?

The next section addresses the materials and methods used. This is followed by the presentation of results and discussion. Finally, the conclusion section is provided.

**Materials and Methods**

To collect review data from the BVA and CNF, two programs were developed using the Python programming language, version 3.10.12. Both programs were executed on a computer running the Linux Mint operating system, version 21.2, with kernel version 5.15.0-
Use of ChatGPT and Tripadvisor data for discovering positive and negative aspects of the BioParque Vale Amazônia and the Carajás National Forest between 2014 and 2022

88-generic, equipped with 11th generation Intel Core i5 processor, 8 gigabytes of RAM, and 512 gigabyte SSD.

The foundation of the first program involves using the Selenium library version 4.14.0 (Selenium, 2024) to access and automatically extract reviews from the Tripadvisor web pages, subsequently storing them in a CSV (comma-separated values) file. The second program utilizes the Requests library version 2.31.0 (Request, 2023) to perform automated queries to ChatGPT 3.5, classifying each review into one of the sentiments: positive, neutral, or negative. The results of each review’s classification were also stored in the CSV file.

The prompt used for ChatGPT to perform sentiment classification was: "Classify the sentiment of the following sentence into just one of the three classes: positive, negative, or neutral. No need to explain the classification, just assign one of the three classes. Sentence: <review text>.”

Subsequently, with the CSV file, the LibreOffice Calc tool (version 7.3.7.2) was used to analyze the data obtained. It is important to note that the following parameters of each review were extracted from Tripadvisor: the city of the user, the state of the user, the number of reviews the user has made on Tripadvisor, the number of likes the review received, the number of stars the user gave to the place (from one to five), the title of the review, the date of the visit, the type of visit (business, solo, friends, couples, or family), the text of the review, the posting date, the place (BVA or CNF), and the sentiment classified by ChatGPT.

Finally, ChatGPT was also consulted to summarize the positive and negative aspects of the BVA and the CNF. For this purpose, the following prompt was used: "Summarize the following text, which contains reviews of a <name of the place>, in no more than 3 topics: <reviews classified as positive and then as negative sentiments>."

In total, 244 reviews were extracted, with 150 from the BVA and 94 from the CNF. The period during which these reviews were posted on Tripadvisor spans from January 2014 to December 2022.

The Figure 1 summarizes the methodology employed in this study.
As mentioned in Section 2, after the extraction of the reviews, data analysis was conducted using LibreOffice Calc tool. Figure 2 displays the percentage of positive, neutral, and negative sentiments for the respective places. It is noticeable that there is a significant predominance of positive sentiment at both the BVA and the CNF.

To confirm ChatGPT’s results about sentiment analysis, Figure 3 shows the number of stars by place. It is possible to check that there are a greater number of comments rated by visitors with five and four stars than with one, two or three stars.
Use of ChatGPT and Tripadvisor data for discovering positive and negative aspects of the BioParque Vale Amazônia and the Carajás National Forest between 2014 and 2022

Figure 3
Percentage of reviews by sentiment and respective places

Through Figure 4, it can be observed that the three most common types of visits at both the BVA and CNF are: with family, with friends, and for business.

Figure 4
Percentage of types of visits by respective places

Another result pertains to the origin of the visitors to the places. Figure 5 shows that the majority of visitors are from the state of Pará itself, with many visitors also coming from Maranhão, Minas Gerais, Rio de Janeiro, and São Paulo.
**Figure 5**

*State of origin of visitors by place*

![State of origin of visitors by place](image)

The summarization, generated by ChatGPT, of the positive and negative aspects of each place is contained in Figure 6.

**Figure 6**

*Summary of positive and negative aspects of the places*

<table>
<thead>
<tr>
<th>Place</th>
<th>Aspect</th>
<th>ChatGPT’s summary</th>
</tr>
</thead>
</table>
| BVA   | Positive | 1. Variety of species;  
2. Attractions and activities;  
3. Accessibility and free admission. |
|       | Negative | 1. Areas without animals;  
2. Bureaucratic authorization process for visits;  
3. Outdated website regarding visitation hours. |
| CNF   | Positive | 1. Immersion in nature;  
2. Diverse attractions;  
3. Emphasis on environmental preservation. |
|       | Negative | 1. Dangerous access road;  
2. Intense mineral exploitation;  
3. Poorly sanitized restaurant and bathrooms. |

Nowacki and Niezgoda (2023) considered data from Tripadvisor and concluded that the appreciation of nature is an important aspect for the good experience of tourists who visited national parks in Poland with family and friends, as happens in BVA and CNF. Conversely, Park et al. (2021) also conducted studies using Tripadvisor data and found that the lack of hygiene in the places displeases visitors, as was perceived in the negative reviews about CNF.
Conclusion

From this study, it was found that 91.33% and 89.36% of the reviews for BVA and CNF, respectively, were classified as positive by ChatGPT. These numbers confirm that the maintenance of these spaces is very well executed.

Furthermore, it is possible to conclude that the research question "what are the positive and negative aspects of BVA and CNF?" was answered in section 3: the positive aspects of both places are similar, emphasizing the well-preserved nature; on the other hand, the negative aspects highlighted regarding the places may have been isolated incidents, as the number of unfavorable reviews is approximately ten times lower than the number of favorable reviews. However, it is always good to optimize communication through the website (in the case of BVA) and ensure proper hygiene of the spaces (in the case of CNF).

This type of analysis allows us to understand the aspects of the places in question that need to be maintained and those that need to be improved, optimizing management. Additionally, it contributes to preserving the image of the places on the internet and in society in general.

In future work, we intend to invest in greater automation of this process, in order to monitor other social networks, make comparisons between periods, and generate alerts for indicators that vary at a determined rate, for example.

References


Received: 04.15.2024  
Accepted: 05.03.2024