






## Introducing an Intelligent Web Scraping (Mining) Method for Analyzing Online Reviews of Tehran Accommodation Users

Athare Ayashi<sup>1</sup>, Ahmad Pourahmad<sup>2</sup>✉, Hassan Ali Faraji Sabokbar<sup>3</sup> , Saeed Zanganeh Shahraki<sup>4</sup> ,  
Fabio Carbone<sup>5</sup> 

1. Department of Human Geograohy and planning, Faculty of Geography, University of Tehran, Tehran

Email: [a\\_ayashi\\_035@ut.ac.ir](mailto:a_ayashi_035@ut.ac.ir)

2. (Corresponding Author) Department of Human Geograohy and planning, Faculty of Geography, University of Tehran, Tehran

Email: [apoura@ut.ac.ir](mailto:apoura@ut.ac.ir)

3. Department of Human Geograohy and planning, Faculty of Geography, University of Tehran, Tehran

Email: [hfaraji@ut.ac.ir](mailto:hfaraji@ut.ac.ir)

4. Department of Human Geograohy and planning, Faculty of Geography, University of Tehran, Tehran

Email: [esoode@yahoo.com](mailto:esoode@yahoo.com)

5. Department of Tourism and Hospitality, University of Northampton, London, England

Email: [fabio.carbone@northampton.ac.uk](mailto:fabio.carbone@northampton.ac.uk)

### ARTICLE INFO

#### Article type:

Research Paper

#### Article History:

Received:

28 October 2024

Received in revised form:

29 January 2025

Accepted:

6 March 2025

Available online:

15 April 2025

#### Keywords:

Web Scraping,

Data Mining,

Online Review Analysis,

Tehran Accommodations.

### ABSTRACT

This research, aimed at analyzing and evaluating online user reviews of accommodations in Tehran, utilizes advanced web scraping techniques. This paper presents a comprehensive automated web scraping method for collecting and analyzing user reviews on an online accommodation platform in Tehran. This smart method includes website selection, tool evaluation, data extraction, preprocessing, and review analysis. By employing this approach, accommodation owners, managers, and marketers can gain deep insights into customer preferences, satisfaction levels, and areas for improvement. Additionally, tourists and users can obtain valuable information from others' experiences when selecting their accommodation. Based on the research findings, the smart web scraping method enables the analysis of large and valuable datasets and can aid in strategic decision-making in the tourism services sector. The findings are presented using descriptive statistics and analytical tests, including t-tests and ANOVA, to assess the mean differences in user reviews across various hotel categories. The study indicates that the average ratings for hotel amenities, room prices, room quality, hotel location, and health protocols are generally above the expected average, reflecting a generally positive perception of Tehran's hotels.

**Citation:** Ayashi, A., Pourahmad, A., Faraji Sabokbar, H. A., Zanganeh Shahraki, S., & Carbone, F. (2025). Introducing an Intelligent Web Scraping (Mining) Method for Analyzing Online Reviews of Tehran Accommodation Users. *Journal of Urban Tourism*, 12 (1), 99-118.

<http://doi.org/10.22059/JUT.2025.378415.1221>



© The Author(s)

This is an open access article under the CC BY NC license (<https://creativecommons.org/licenses/by/4.0/>).

**Publisher:** University of Tehran Press

## Extended Abstract

### Introduction

In the age of digital information and online platforms, the tourism and hospitality industry has witnessed significant changes in how customers share their opinions and experiences. Nowadays, making informed decisions about travel, destinations, and accommodations heavily relies on online reviews by tourists. These reviews have become essential for travelers seeking the best travel and accommodation experiences. The manual analysis becomes impractical because of the importance of these opinions and experiences in tourists' decision-making processes and the vast volume of reviews. As a result, smart web scraping methods have been employed on online platforms for tourism accommodations. Web scraping is developing a computer program to automatically download, analyze, and organize data from web pages, making it highly practical for extracting data from multiple pages simultaneously. The abundance of general tourism data available online holds significant data analysis potential. However, much of this data remains unanalyzed and underutilized. By collecting and analyzing these untapped data sources, significant improvements can be made in the tourism sector of any destination.

As mentioned, the topic of online customer reviews on tourism and hospitality websites and platforms is highly significant and has been emphasized in previous studies. While this topic has been extensively addressed in international research, it has received far less attention in Iran, especially for accommodations and hotels. Therefore, this study refers to the data collection method through web mining and reviews its importance and key process stages. Furthermore, in order to demonstrate the significance of the data collected via web scraping, descriptive statistics, differences in mean values of online user reviews, and significant differences in reviews for star-rated accommodations are calculated. Additionally, the reasoning behind choosing the website "Eghamat 24" for web scraping and analysis of online user reviews is discussed.

### Methodology

This study aims to uncover hidden opinions in reviews and user feedback on the accommodations of accommodations in Tehran and determine if there are differences among accommodations based on review scores. Consequently, this study involves extracting user reviews from online review platforms. The extracted reviews are analyzed and converted into documents, tables, and graphs. Further analysis focuses on uncovering hidden opinions within the reviews, and scores are calculated from the feedback. This paper examines the steps and processes involved in web scraping, which include:

1. Website analysis;
2. Web scraping;
3. Data extraction;
4. Organizing, processing, and storing the data.

The study utilizes Selenium for web scraping. Selenium is a powerful and popular tool that provides a framework for automating web browsers. Essentially, it is used for scraping dynamic web pages, allowing interaction with web pages such as clicking buttons or filling out forms and extracting data from websites. The first step is installing Selenium, for which the corresponding code is written in Python. The next step involves identifying the URLs of the profiles for each accommodation, which is done using the WebDriver command from the Selenium library. After that, the URLs are accessed and clicked automatically. The next stage involves navigating the tags and elements within the URLs using the XPath command. Finally, after identifying all the tags and classes within the website's HTML source code, a for loop is used to extract all the tags and elements across all the pages. The Pandas library in Python is employed to save the data in an Excel file.

### Results and discussion

The findings of the study, based on descriptive statistics and analytical tests, show that the average ratings for various aspects of hotels, such as services, room price, quality, location, and cleanliness, are generally above the expected average, indicating an overall positive perception of

hotels in Tehran. The paper highlights the potential of web scraping as a powerful technique for automatically collecting data from websites, which can significantly contribute to enhancing tourism services and making strategic decisions in the industry. By studying and analyzing these reviews, businesses can better understand tourists' needs and preferences and implement necessary improvements in accommodations. Some of the impacts of using automated web scraping for businesses and users include:

1. Improving service quality;
2. Better decision-making by users;
3. Marketing strategy enhancements;
4. Boosting business trust.

### **Conclusion**

The impact of online reviews and customer feedback from tourists is becoming increasingly significant. Tourists trust online reviews when planning their trips, while businesses use them to create effective marketing strategies. However, analyzing individual reviews manually is impractical due to the sheer volume of available data. As a result, considerable efforts have been made

in recent years to develop methods for automatic analysis and summarization of reviews. This research demonstrates how various analyses conducted on user feedback data from online platforms concerning accommodations in Tehran can help businesses and users, ultimately leading to improved business practices.

### **Funding**

There is no funding support.

### **Authors' Contribution**

Authors contributed equally to the conceptualization and writing of the article. All of the authors approved the content of the manuscript and agreed on all aspects of the work declaration of competing interest none.

### **Conflict of Interest**

Authors declared no conflict of interest.

### **Acknowledgments**

We are grateful to all the scientific consultants of this paper.

## ارائه روش هوشمند وب کاوی (خزش) برای تحلیل نظرات برخط استفاده‌کنندگان از اقامتگاه‌های گردشگری شهر تهران

اطهره عیاشی<sup>۱</sup>، احمد پوراحمد<sup>۲</sup>✉، حسنعلی فرجی سبکبار<sup>۳</sup>، سعید زنگنه شهرکی<sup>۴</sup>، فابیو کاربونه<sup>۵</sup>

- ۱- گروه جغرافیای انسانی و برنامه‌ریزی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران. رایانامه: [a\\_ayashi\\_035@ut.ac.ir](mailto:a_ayashi_035@ut.ac.ir)
- ۲- نویسنده مسئول، گروه جغرافیای انسانی و برنامه‌ریزی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران. رایانامه: [apoura@ut.ac.ir](mailto:apoura@ut.ac.ir)
- ۳- گروه جغرافیای انسانی و برنامه‌ریزی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران. رایانامه: [hfaraji@ut.ac.ir](mailto:hfaraji@ut.ac.ir)
- ۴- گروه جغرافیای انسانی و برنامه‌ریزی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران. رایانامه: [saeed.zanganeh@ut.ac.ir](mailto:saeed.zanganeh@ut.ac.ir)
- ۵- گروه گردشگری و هتلداری، دانشگاه نورث همپتون، لندن، انگلستان. رایانامه: [fabio.carbone@northampton.ac.uk](mailto:fabio.carbone@northampton.ac.uk)

### اطلاعات مقاله

### چکیده

#### نوع مقاله:

مقاله پژوهشی

#### تاریخ دریافت:

۱۴۰۳/۰۸/۰۷

#### تاریخ بازنگری:

۱۴۰۳/۱۱/۱۰

#### تاریخ پذیرش:

۱۴۰۳/۱۲/۱۶

#### تاریخ چاپ:

۱۴۰۴/۰۱/۲۶

#### واژگان کلیدی:

وب کاوی،  
خزنده نویسی،  
تحلیل آنلاین نظرات،  
اقامتگاه‌های تهران.

این پژوهش، با هدف تحلیل و ارزیابی نظرات برخط کاربران درباره اقامتگاه‌های شهر تهران، از تکنیک‌های پیشرفته وب کاوی (Web Scraping) استفاده کرده است. از این رو در این مقاله، به بررسی یک روش جامع خودکار وب کاوی برای جمع‌آوری و تجزیه و تحلیل نظرات کاربران در یکی از سایت‌های آنلاین اقامتگاه در شهر تهران ارائه شده است. این روش هوشمند شامل مراحل مانند انتخاب وب‌سایت، بررسی ابزارها، استخراج داده‌ها، پیش‌پردازش و تجزیه و تحلیل نظرات است. با استفاده از این روش، صاحبان اقامتگاه‌ها، مدیران و بازاریابان می‌توانند بینشی عمیق درباره ترجیحات مشتریان، سطح رضایت و نیازهای بهبود را به دست آورند. همچنین گردشگران و کاربران برای انتخاب اقامتگاه خود می‌توانند اطلاعات ارزشمندی از تجربیات سایرین کسب نمایند. بر اساس نتایج تحقیق، روش هوشمند وب کاوی، امکان تحلیل داده‌های بزرگ و ارزشمند را فراهم می‌کند و می‌تواند به تصمیم‌گیری‌های استراتژیک در حوزه خدمات گردشگری کمک کند. یافته‌ها با استفاده از آمار توصیفی و آزمون‌های تحلیلی شامل آزمون‌های t و ANOVA، برای ارزیابی میانگین تفاوت در نظرات کاربران در دسته‌های مختلف هتل ارائه شده‌اند. این مطالعه نشان می‌دهد که میانگین امتیازات امکانات هتل، قیمت اتاق، کیفیت اتاق، موقعیت هتل و پروتکل‌های بهداشتی عموماً بالاتر از میانگین مورد انتظار است که نشان‌دهنده درک کلی مثبت از هتل‌های تهران است.

**استناد:** عیاشی، اطهره؛ پوراحمد، احمد؛ فرجی سبکبار، حسنعلی؛ زنگنه شهرکی، سعید و کاربونه، فابیو. (۱۴۰۴). ارائه روش هوشمند وب کاوی (خزش) برای تحلیل نظرات برخط استفاده‌کنندگان از اقامتگاه‌های گردشگری شهر تهران. *مجله گردشگری شهری*، ۱۱۲(۱)، ۹۹-۱۱۸.

<http://doi.org/10.22059/JUT.2025.378415.1221>

## مقدمه

در عصر اطلاعات دیجیتال و پلتفرم‌های آنلاین، صنعت گردشگری و مهمان‌نوازی شاهد تغییرات قابل توجهی در بیان نظرات و اشتراک‌گذاری تجربیات مشتریان بوده است (Choong, 2019: 2; Pang, 2023: 19; Han & Anderson, 2021: 89). به طوری که تصمیم‌گیری آگاهانه انتخاب سفر، مقصد و اقامت به شدت به بررسی‌های آنلاین توسط گردشگران متکی است، هم‌چنین بررسی‌های آنلاین به یک منبع ضروری برای گردشگرانی تبدیل شده است که به دنبال تجربه سفر و اقامتگاه مناسب هستند (Yildirim et al, 2023: 138; Renganathan & Upadhy, 2021: 166; Rajan & Shyam, 2015: 1). این بررسی‌ها و نظرات، بینش ارزشمندی را در مورد تجربیات کاربران ارائه می‌دهند و جنبه‌هایی مانند تمیزی، خدمات، موقعیت مکانی و قیمت را به تفصیل بیان می‌کنند (Fazzolari et al, 2018: 38; Adhinugroho et al, 2020: 8; Nwakanma et al, 2019: 60; Choong, 2019: 3). با این حال، با حجم گسترده‌ای از اطلاعات موجود در وبسایت‌های متعدد رزرو هتل و اقامتگاه، تجزیه و تحلیل دستی این بررسی‌ها به یک کار خسته‌کننده و وقت‌گیر تبدیل می‌شود و به طور دستی امکان‌پذیر نمی‌باشد (Nwakanma et al, 2019: 60; Alaei et al, 2019: 175; Gheorghie et al, 2018: 63). از طرفی استخراج و تجزیه و تحلیل این داده‌ها می‌تواند به صاحبان اقامتگاه‌ها، مدیران و بازاریابان بینش مهمی در مورد ترجیحات مشتری، سطح رضایت و زمینه‌های بهبود ارائه دهد (Beck et al, 2017: 40; Barbera et al, 2023: 2).

وب کاوی، به عنوان خزش در وب شناخته می‌شود و به عنوان یک تکنیک قدرتمند برای جمع‌آوری حجم زیادی از داده‌ها از وبسایت‌ها به صورت خودکار ظاهر شده است (Barbera et al, 2023: 2). با استفاده از وب کاوی، پژوهشگران و کسب‌وکارها می‌توانند حجم زیادی از محتوای تولیدشده توسط کاربر، مانند بررسی هتل‌ها، رتبه‌بندی و نظرات را از پلتفرم‌های آنلاین مختلف جمع‌آوری کنند (Dogucu & Rundel, 2021: 112; Fazzolari & Petrocchi, 2018: 165; Stringam et al, 2023: 56-57; Yuan, 2023: 25; Renganathan & Upadhy, 2021: 165). می‌توان این داده‌ها را برای کشف الگوها، احساسات و اطلاعات عملی که می‌تواند تصمیم‌گیری استراتژیک را هدایت کند، تجزیه و تحلیل کرد (Sahu et al, 2022: 28; Renganathan & Upadhy, 2021: 169).

در این راستا، تکنیک وب کاوی یک روش قدرتمند برای جمع‌آوری و تحلیل خودکار حجم بزرگی از داده‌های پلتفرم‌ها و وبسایت‌هاست که به پژوهشگران و کسب‌وکارهای امکان می‌دهد تا حجم زیادی از محتوای تولیدشده توسط کاربران را از پلتفرم‌های آنلاین مختلف استخراج و تحلیل کنند. هم‌چنین از آنجایی که استخراج داده‌ها به صورت دستی نه تنها زمان است، بلکه ممکن است دقت و جامعیت لازم را نیز نداشته باشد. از این رو، یک روش هوشمند وب کاوی برای استخراج، تحلیل و طبقه‌بندی این نظرات وجود دارد تا مدیران اقامتگاه‌ها و گردشگران بتوانند تصمیمات آگاهانه‌تری بگیرند. نظرات آنلاین به عنوان یک موضوع پژوهشی در مقاصد ایران بدان کمتر توجه شده است در حالی که پلتفرم‌های آنلاین در ایران نیز وجود دارد. بنابراین، مسئله اصلی تحقیق این است که چگونه می‌توان از روش‌های هوشمند وب کاوی برای تحلیل نظرات برخط استفاده‌کنندگان از اقامتگاه‌های شهر تهران بهره برد تا اطلاعات ارزشمندی درباره ترجیحات مشتری، سطح رضایت و زمینه‌های بهبود خدمات به دست آورد؟ بنابراین، اهمیت و ضرورت این پژوهش در توانایی آن برای ارائه بینش‌های عمیق و کاربردی از نظرات کاربران آنلاین نهفته است، که می‌تواند به بهبود مدیریت و ارتقاء کیفیت خدمات در صنعت گردشگری منجر شود. با توجه به رقابت شدید در این صنعت، بهره‌گیری از تکنیک‌های وب کاوی برای تحلیل داده‌های تولیدشده توسط کاربران، ابزاری حیاتی برای دستیابی به موفقیت و حفظ رضایت مشتریان به شمار می‌آید. از این رو، هدف این مقاله، ارائه یک روش جامع وب کاوی و ایجاد پایه‌های محکم برای جمع‌آوری داده‌های گردشگری است که

به طور خاص برای تجزیه و تحلیل نظرات کاربران اقامتگاه‌های شهر تهران طراحی شده است. هم‌چنین در خصوص مراحل کلیدی فرایند وب کاوی از جمله استخراج داده‌ها، پیش‌پردازش، تجزیه و تحلیل نظرات را بررسی خواهیم کرد.

## مبانی نظری

وب‌سایت‌های آنلاین فرصت‌های جذابی را برای پژوهشگران حوزه گردشگری و مهمان‌نوازی ایجاد کرده‌اند. به طوری که هر وب‌سایت یک منبع داده بالقوه محسوب می‌شود. در این وب‌سایت‌ها می‌توان نظرات کاربران را جمع‌آوری کرد و به تجزیه و تحلیل آن‌ها پرداخت. رضایتمندی آن‌ها را جستجو کرد یا رابطه بین قیمت و درجه‌بندی آن‌ها را تعیین نمود (Han & Anderson, 2021: 1). از این‌رو، سیستم‌های اطلاعاتی موجود در اینترنت، داده‌های ارزشمندی را در مورد اطلاعات گردشگری و رفتارهای گردشگران در اختیار صنعت گردشگری قرار می‌دهد. به‌عنوان مثال وب‌سایت‌هایی مانند Hotel.com، Agoda و Tripadvisor بهترین توصیه‌ها را درباره مکان‌ها، اتاق‌ها و محدوده‌های قیمتی ارائه می‌دهند. کاربران این پلتفرم با به اشتراک گذاشتن تجربیات خود، ارائه امتیازات به مکان‌ها، و اظهار نظر در مورد جاذبه‌های گردشگری مختلف در سراسر جهان درگیر می‌شوند. این بررسی‌ها می‌توانند بر تصمیم‌گیری گردشگران در برنامه‌ریزی سفرهایشان تأثیر بگذارند (Kamarazaman et al., 2024: 16؛ حق وردی زاده دهلیق و همکاران، ۱۴۰۲: ۱۳۱-۱۳۲).

داده‌های گردشگری عمومی زیادی که در اینترنت موجود است، می‌توانند دارایی‌های بالقوه ارزشمندی برای تحلیل داده‌ها باشند، اما بیشتر آن‌ها بدون تحلیل رها شده‌اند. این داده‌های هدر رفته می‌توانند جمع‌آوری شده و برای بهبود حوزه گردشگری هر مقصدی استفاده شوند (Choong, 2019: 1؛ میمندی و همکاران، ۱۴۰۲: ۱۸).

یکی از ذی‌نفعان اصلی مدیریت مقصد، صنعت هتلداری است (Oses Fernández et al., 2018: 1). تحلیل هتل‌ها یک روند پژوهشی جدید است که از تحلیل داده‌ها به وجود آمده و هدف آن بهره‌برداری از دانش گردشگری بر اساس منابع موجود از داده‌های بزرگ (Big Data) است. یکی از برجسته‌ترین این منابع، شبکه‌های اجتماعی و پلتفرم‌های گردشگری است (Moro et al., 2019a: 1113; Moro et al., 2018: 3; Aluri et al., 2019: 79). صنعت گردشگری و هتلداری یکی از پیشگامان در پیاده‌سازی پلتفرم‌های محبوب بررسی آنلاین در سطح جهان است. یکی از شناخته‌شده‌ترین این پلتفرم‌ها، تریپ ادوایزر است که در آن کاربران تشویق می‌شوند تا نظرات خود را درباره هتل‌هایی که در آن اقامت داشته‌اند بنویسند (Moro et al., 2019b: 88; Moro et al., 2018: 4). نویسندگان مختلف از این پلتفرم‌ها و داده‌های موجود برای تحقیقات مختلفی استفاده کرده‌اند. به‌عنوان مثال، رادویویچ<sup>۱</sup> و همکاران (2015) با استفاده از سایت booking.com به بررسی تأثیر ویژگی‌های هتل بر رضایت مشتریان در ۴۷ پایتخت اروپایی پرداخته است. آن‌ها ویژگی‌های هتل مانند رتبه‌بندی ستاره‌ای (از ۱ تا ۵)، برند هتل، مکان هتل، اندازه هتل (تعداد کل اتاق‌ها) قیمت و امکانات هتل را مورد بررسی قرار دادند. نویسندگان نتایج را از دیدگاه نظریه همسانی-تضاد توضیح دادند و نتیجه‌گیری کردند که زمانی که انتظارات مشتری و تجربه درک شده از هتل در سطح پذیرش خاصی قرار دارد، نظرات مشتریان منعکس‌کننده رتبه‌بندی ستاره‌ای هتل است. در مطالعه دیگری لیو<sup>۲</sup> و همکاران (2017)، با توجه به شواهد تجربی که نشان‌دهنده تأثیر ملیت مشتریان هتل بر رضایت کلی و درک آن‌ها از اهمیت ویژگی‌های مختلف هتل است، امتیازات رتبه‌بندی تریپ ادوایزر برای چهار ویژگی هتل شامل تمیزی و نظافت، مکان، اتاق ارزش و امکانات هتل (خدمات) را برای پنج هتل در چین میان

1. Radojevic

2. Liu

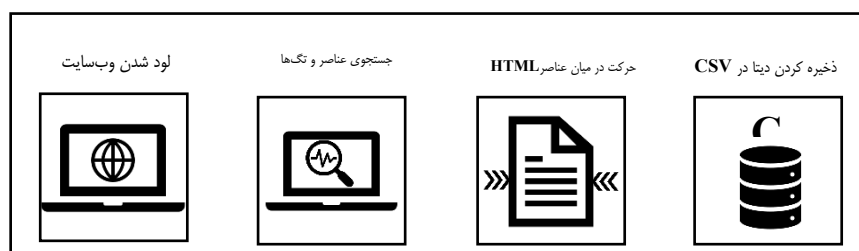
مشتریان ژاپنی، چینی و روسی مقایسه کردند. نتایج نشان داد که ملیت نقش مهمی در تعیین امتیازات مشتریان برای ویژگی‌های انتخاب‌شده ایفا می‌کند. مشتریان چینی، مکان و خدمات را نسبتاً پایین ارزیابی کردند، درحالی‌که اتاق‌های هتل را بسیار بالاتر ارزیابی کردند. مشتریان روس اتاق‌ها را پایین‌تر ارزیابی کردند اما ارزش را بالاتر درک کردند. مشتریان ژاپنی امتیازات بالاتری برای مکان و خدمات در مقایسه با سایر ویژگی‌ها ارائه دادند. برنزینا<sup>۱</sup> و همکارانش (۲۰۱۶) به بررسی تفاوت‌های میان مشتریان راضی و ناراضی هتل‌ها پرداخته‌اند. در این تحقیق از روش تحلیل متن برای استخراج اطلاعات از بررسی‌های آنلاین هتل‌های استفاده کرده است. از آنجایی‌که هتل‌ها در یک محیط رقابتی و پویای کسب‌وکار فعالیت می‌کنند و برای موفقیت، نیازمند درک عمیق‌تری از تجربیات مشتریان خود هستند، رضایت مشتریان عامل مهمی در تضمین بازگشت مجدد آن‌ها و توصیه هتل به دیگران است. درحالی‌که ناراضی می‌تواند منجر به تبلیغات منفی شود. داده‌های این تحقیق از بررسی‌های آنلاین سایت تریپ ادوایزر جمع‌آوری شده است و با استفاده از تحلیل متن، به بررسی‌های مثبت و منفی مشتریان پرداخته شده است. نتایج تحقیق نشان داد که مشتریان راضی بیشتر به جنبه‌های ناملموس اقامت خود مانند رفتار کارکنان اشاره می‌کنند، درحالی‌که مشتریان ناراضی بیشتر بر جنبه‌های ملموس مانند اثاثیه و مسائل مالی تمرکز دارند. این تفاوت در تمرکز مشتریان نشان‌دهنده اهمیت رفتار کارکنان و خدمات در ایجاد رضایت مشتریان است. وانگ<sup>۲</sup> و همکاران (۲۰۱۹) از طریق پلتفرم آنلاین تریپ ادوایزر و با استفاده از داده‌های گسترده شامل نظرات و امتیازات مشتریان هتل‌ها، به اهمیت کیفیت خدمات در صنعت هتلداری پرداخته است و نشان می‌دهد که چگونه ارزش ادراک‌شده و رتبه‌بندی ستاره‌ای هتل‌ها می‌توانند بر رضایت مشتریان تأثیر بگذارند. از این‌رو، به بررسی عواملی پرداخته شده است که بر رضایت مشتریان هتل‌ها اثر می‌گذارد. بر اساس نتایج محققان تلاش کرده‌اند تا نشان دهند که هرچه کیفیت خدمات ارائه‌شده توسط هتل بالاتر باشد، رضایت مشتریان نیز بیشتر خواهد بود. از این‌رو، رتبه‌بندی ستاره‌ای هتل‌ها نقش تعدیلی در رابطه بین کیفیت خدمات و رضایت ایفا می‌کند. به طوری‌که در هتل‌های سه ستاره، کیفیت خدمات اهمیت بیشتری دارد درحالی‌که در هتل‌های ۴ و ۵ ستاره عوامل دیگری نظیر امکانات و خدمات لوکس‌تر اهمیت بیشتری دارند. فیلیپس<sup>۳</sup> و همکارانش (۲۰۱۷) نیز بر تأثیر نظرات آنلاین به‌ویژه نظرات مثبت و منفی بر عملکرد هتل‌ها تأکید کرده‌اند. محققان با استفاده از داده‌های جمع‌آوری‌شده از ۶۸ پلتفرم آنلاین و ۴۲ هتل در سوئیس جنبه‌های فیزیکی، کیفیت غذا و نوشیدنی و جنبه‌های انسانی ارائه خدمات را برای ارزیابی عملکرد هتل‌ها از شاخص‌هایی مانند درآمد به ازای هر اتاق قابل‌دسترس و نرخ اشغال اتاق‌ها استفاده کرده‌اند. نتایج نشان می‌دهد که ویژگی هتل‌ها از جمله کیفیت اتاق‌ها، ارائه اینترنت و ساختمان هتل، بیشترین تأثیر را بر عملکرد هتل دارند و نظرات مثبت، بیشترین تأثیر را بر تقاضای مشتریان دارند. احمد و سان<sup>۴</sup> (۲۰۱۸) نیز به تحلیل عواملی که موجب بی‌اعتمادی مصرف‌کنندگان به نظرات هتل‌های آنلاین می‌شوند، می‌پردازد. از طرفی متذکر می‌شوند که نظرات آنلاین هتل‌ها به دلیل ناشناس بودن به‌راحتی می‌توانند اطلاعات گمراه‌کننده را منتشر کنند. بی‌اعتمادی موجب ناراحتی روانی مصرف‌کنندگان و در نهایت به رفتارهای منفی مانند کاهش خرید مجدد و افزایش کلام منفی الکترونیکی منجر می‌شود. نتایج مطالعه می‌تواند به وبسایت‌های مسافرتی آنلاین و شرکت‌های هتلداری کمک کند تا استراتژی‌هایی برای افزایش اعتماد و بازگرداندن مشتریان ناراضی توسعه دهند. با توجه به در دسترس بودن آمار گردشگری به‌طور گسترده در اینترنت، می‌توان اطلاعات مربوط به خدمات گردشگری را جمع‌آوری کرد و برای تولید آمار استفاده کرد و شرایط و پدیده یک منطقه را تشریح کرد (Adhinugroho et al., 2020: 2)

1. Brenzina
2. Wong
3. Phillips
4. Ahmad & Sun

همان‌طور که اشاره شد، موضوع نظرات آنلاین مشتریان در وب‌سایت‌ها و پلتفرم‌های گردشگری و هتلداری دارای اهمیت است که در تحقیقات گذشته نیز بر آن تأکید شده است و از آن‌ها استفاده کرده‌اند. آنچه مشخص است، در مقالات خارجی به این موضوع پرداخته‌اند اما در ایران و برای اقامتگاه‌ها و هتل‌های ایران کمتر به آن توجه شده است. در ایران نیز، پلتفرم‌های آنلاین برای اقامتگاه‌های ایرانی نیز وجود دارد که از جمله می‌توان به اسنپ تریپ، اقامت ۲۴، ایران هتل و هتل یار اشاره کرد. بنابراین، در این پژوهش به تحلیل آماری داده‌های امتیازبندی شده سایت اقامت ۲۴ در شهر تهران به همراه ارائه یک روش جامع وب کاوی و ایجاد پایه‌های محکم برای جمع‌آوری داده‌های گردشگری پرداخته شده است. آنچه در پژوهش‌های پیشین مشاهده شده است این است که داده‌ها به صورت دستی از سایت‌ها و پلتفرم‌های برخط گردآوری شده‌اند و به روش استخراج خودکار داده‌ها کمتر پرداخته‌اند. از این رو، در این پژوهش به روش جمع‌آوری از طریق وب کاوی (web scraping) اشاره شده است و اهمیت و مراحل کلیدی فرایند نیز بررسی می‌شود. پس از آن به منظور نمایش اهمیت داده‌های جمع‌آوری شده از طریق وب کاوی، آمارهای توصیفی داده‌ها، تفاوت میانگین نظرات برخط استفاده‌کنندگان و همچنین تفاوت معنی‌داری نظرات برای رتبه‌بندی‌های ستاره‌ای اقامتگاه‌ها محاسبه می‌شود. همچنین دلیل انتخاب سایت اقامت ۲۴ برای وب کاوی و تحلیل نظرات آنلاین کاربران،

### وب کاوی

وب کاوی، فرایندی است که از ایجاد یک برنامه کامپیوتری برای دانلود، تجزیه و تحلیل و سازمان‌دهی داده‌ها از صفحات وب به روشی خودکار به وجود می‌آید (Sahu et al, 2022: 25; Han & Anderson, 2021: 89) و برای استخراج داده از چندین صفحه به صورت هم‌زمان بسیار کاربردی است (Luo et al, 2020: 86). این روش برای استخراج داده‌های مفید و مرتبط مانند متن، تصاویر، پیوندها و داده‌های ساختاریافته از صفحات HTML است و از زبان برنامه‌نویسی برای اتصال به صفحات سروری استفاده می‌کند (Sequeira et al, 2020: 2). به طور کل تکنیک‌های وب کاوی به کاربران اجازه می‌دهد تا اطلاعات بدون ساختار را از اینترنت جمع کنند و آن را به شکلی ساختاریافته تبدیل کنند (Dogucu & Rundel, 2021: 112). داده‌ها از سایت‌ها با استفاده از پروتکل HTTP مورد استفاده مرورگرهای وب استخراج می‌شوند. این فرایند را می‌توان با مرورگرهای دستی یا خودکار با خزنده‌های وب انجام داد. وب کاوی یکی از بارزترین ابزارهای موجود برای پژوهشگران حوزه داده است زیرا امکان استخراج حجم عظیمی از داده‌هایی را که دائماً به صورت آنلاین تولید می‌شوند را با هزینه نسبتاً کم فراهم می‌کند (khder, 2021: 183; Pokhrel et al, 2023: 302).



شکل ۱. مراحل وب کاوی، منبع: Pokhrel et al, 2023: 302



۱- خزش وب<sup>۱</sup>: خزش در وب با وب کاوی<sup>۲</sup> شروع می‌شود که شامل پیمایش در وبسایتها و جمع‌آوری URL ها و محتوای صفحه وب است. این فرآیند معمولاً از خزنده‌های وب یا عنکبوت‌هایی استفاده می‌کند که به‌طور سیستماتیک وب را طی می‌کنند، پیوندها را دنبال می‌کنند و صفحات وب را فهرست می‌کنند.

۲- تجزیه HTML<sup>۳</sup>: هنگامی که صفحات وب موردنظر شناسایی شدند، ساختار HTML (یا XML) صفحات برای استخراج داده‌های مربوطه تجزیه می‌شود. تجزیه شامل مکان‌یابی و استخراج عناصر و تگ‌های خاص HTML مانند سرفصل‌ها، پاراگراف‌ها، جداول یا تصاویر است.

۳- استخراج داده‌ها: پس از تجزیه ساختار HTML، داده‌های موردنظر با استفاده از تکنیک‌های مختلف استخراج می‌شود. این ممکن است شامل عبارات منظم، XPath، انتخاب‌گرهای CSS یا روش‌های پیشرفته‌تر مانند استفاده از API یا مرورگرهای بدون سر باشد.

۴- پردازش و ذخیره‌سازی داده‌ها: داده‌های استخراج‌شده اغلب به پردازش، تمیز کردن و تبدیل بیشتر نیاز دارند تا برای تجزیه و تحلیل آماده شوند. این ممکن است شامل حذف موارد تکراری، مدیریت مقادیر از دست‌رفته، و ساختار داده‌های بدون ساختار باشد. سپس داده‌های پردازش‌شده را می‌توان در یک پایگاه داده ذخیره کرد یا به فرمت مناسب برای تجزیه و تحلیل خروجی گرفت (Gheorghie et al, 2018: 70; Ali et al, 2022: 3; Han & Anderson, 2021: 92 – 98).

بر اساس فرایند و مراحل وب کاوی، خدر<sup>۴</sup> (۲۰۲۱)، دانش فنی موردنیاز در هر مرحله را در قالب شکل (۳) تشریح می‌کند:



شکل ۲. فرایند وب کاوی



شکل ۳. ملاحظات فنی و دانش موردنیاز وب کاوی، منبع: khder, 2021: 184

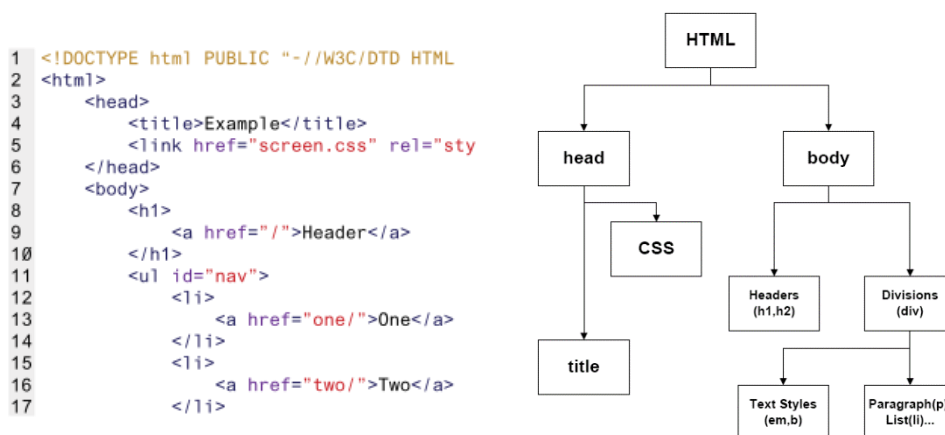
### ابزارهای فنی وب کاوی

❖ پیمایش تگ‌های HTML و برگه‌های آبخاری:

HTML و CSS زبان استاندارد است که در پشت اکثر هر سند وب قرار دارد. کاربران وب با HTML تعامل دارند و هدف اصلی آن رسیدگی به محتوای یک صفحه وب مانند متن، لینک، تصاویر و ... است. برگه‌های آبخاری (CSS) یکی

1. Web Crawling
2. Web Scraping
3. HTML Parsing
4. Khder

از فناوری‌های اصلی برای ساخت وب است. HTML برای ساختار صفحه وب کاربرد دارد و CSS استایل و طرح‌بندی را ارائه می‌کند (Dogucu & Rundel, 2021: 113). ساختار HTML و CSS به‌طور مختصر در شکل ۳ آورده شده است:



شکل ۴. ساختار صفحات HTML

### جعبه‌ابزار سلنیوم<sup>۱</sup>

از تکنولوژی وب کاوی (خزنده) پایتون می‌توان برای خودکار سازی فرایند جمع‌آوری داده‌ها استفاده کرد (Yuan, 2023: 25). ابزارها و کتابخانه‌های Beautiful Soup, Scrapy, Selenium و request در پایتون با ارائه قابلیت‌هایی برای پیمایش در صفحات وب، مکان‌یابی عناصر خاص و استخراج داده‌های مرتبط به‌طور مؤثر، فرایند وب کاوی را ساده می‌کنند. این کتابخانه‌ها قابلیت‌هایی را برای تجزیه HTML، تعامل با صفحات وب و استخراج داده‌ها فراهم می‌کنند. (Sharma & Borkar, 2023: 243). در میان کتابخانه‌های بالا، سلنیوم یک ابزار قدرتمند و محبوب وب کاوی است که چارچوبی برای خودکار سازی مرورگرهای وب فراهم می‌کند. در واقع سلنیوم ابزاری برای وب کاوی صفحات پویا است و این امکان را می‌دهد که با صفحات وب تعامل ایجاد شود و اقداماتی نظیر کلیک کردن روی دکمه‌ها و یا پر کردن فرم‌ها و استخراج داده از وب‌سایت‌ها انجام پذیرد (Sequeira et al, 2020: 3).

### روش پژوهش

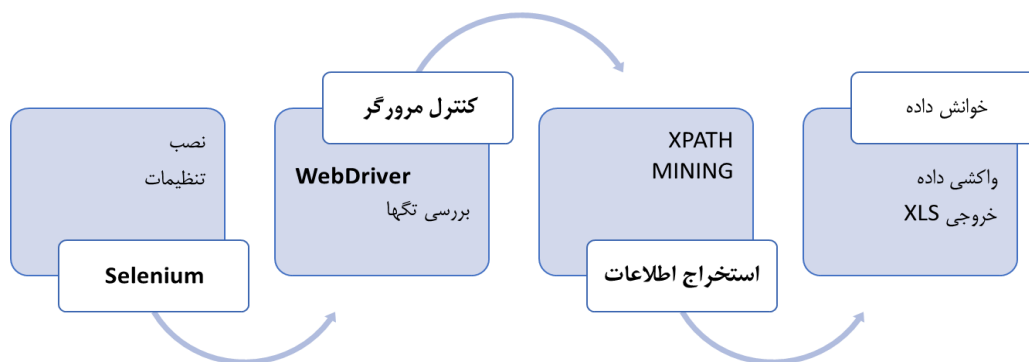
هدف این مطالعه یافتن نظرات پنهان در بررسی‌ها و نظرات کاربران در مورد خدمات اقامتگاه‌های شهر تهران و یافتن اینکه آیا تفاوت‌های بین اقامتگاه‌ها بر اساس امتیازات نظرات می‌باشد. بنابراین این مطالعه شامل استخراج نظرات کاربران از وب‌سایت‌های آنلاین بررسی نظرات است. بررسی‌های استخراج‌شده تجزیه و تحلیل می‌شود و به اسناد، جداول و گراف‌ها تبدیل می‌شوند. تجزیه و تحلیل‌ها بیشتر برای یافتن نظرات پنهان در بررسی‌ها است و امتیازات حاصل از نظرات محاسبه می‌شود. این مقاله شامل ابزارهای وب کاوی است. بدین منظور، از میان سایت‌های آنلاین بررسی نظرات اقامتگاه‌ها، سایت eghamat24 انتخاب شد. دلیل اصلی انتخاب این سایت، رعایت استانداردهای طراحی وب‌سایت و بر این اساس قابلیت وب کاوی آن، ارائه لیست کاملی از اقامتگاه‌های شهر تهران در آن می‌باشد.

نظرات کاربران در مورد اقامتگاه‌های شهر تهران با استفاده از تکنیک وب کاوی از سایت اقامت ۲۴ به آدرس <https://www.eghamat24.com/TehranHotels.html> استخراج شد. پس از بررسی‌های اولیه از پروفایل هر کدام از اقامتگاه‌ها، عناصر و تگ‌های هر کدام از شاخص‌های امتیازات استخراج شد.



شکل ۵. استخراج و بررسی تگ‌های سایت، منبع: سایت اقامت ۲۴

به‌منظور وب کاوی در اقامت ۲۴، ابتدا از زبان برنامه‌نویسی پایتون و کتابخانه سلنیوم استفاده شد تا نظرات کاربران اقامتگاه‌های شهر تهران از استخراج شود. جمع‌آوری نظرات بر اساس فیلدهای مدنظر (شامل امکانات هتل، قیمت، کیفیت اتاق، موقعیت جغرافیایی و پروتکل‌های بهداشتی) مورد بررسی قرار گرفت. مراحل و فرایند انجام وب کاوی با سلنیوم در بستر زبان برنامه‌نویسی پایتون به شرح شکل (۶) می‌باشد:



شکل ۶. فرایند وب کاوی با سلنیوم

```

4 # کتابخانه سلنیوم
5 from selenium import webdriver
6 from selenium.webdriver.common.keys import Keys
7 from selenium.webdriver.common.by import By
  
```

شکل ۷. نصب سلنیوم در پایتون

مرحله اول نصب سلنیوم می‌باشد که کدهای مربوط به آن به شکل زیر در محیط پایتون نوشته می‌شود: پس از آن، پیدا کردن URL های پروفایل هریک از اقامتگاه‌های می‌باشد که با استفاده از دستور WebDriver از کتابخانه سلنیوم این اقدام انجام می‌شود و پس از آن خودکار کردن کلیک روی هر کدام از URL ها است. مرحله بعد پیدا کردن و حرکت در میان تگ‌ها و عناصر داخل URL هاست که با دستور XPath انجام می‌پذیرد.

**WebDriver**: این دستور برای اکثر مرورگرهای اینترنتی وجود دارد. هنگام اجرای اسکریپت، یک پنجره مرورگر باز می‌شود و اقدامات مشخص شده و کد نویسی شده را به صورت خودکار انجام می‌دهد (Ali et al: 2022: 3). در این مطالعه از مرورگر گوگل کروم برای دستور WebDriver استفاده شد.

```

16 باز کردن وب سایت
17 driver = webdriver.Chrome()
18 driver.get('https://www.eghamat24.com/TehranHotels.html')

```

شکل ۸. کنترل مرورگر در سلیوم

XPath: هنگامی که WebDriver صفحه وب را باز می‌کند، هدف استخراج محتوای نمایش داده شده در صفحه و ذخیره آن است. از این رو، باید کد منبع (Source Code) را از بخش Inspect مرورگر استخراج و وب کاوی کرد (Ali et al: 2022: 3). در این مطالعه از روش XPath برای استخراج تگ‌ها و ویژگی‌های تگ‌ها استفاده شد.

```

33 پیدا کردن عناصر اقامتگاه ها
34 hotel_elements = driver.find_elements(by=By.XPATH, value="//a[@class = 'd-block mb-2']")
35
36 استخراج تگ ها و عناصر
37 Hotels_urls = [link.get_attribute("href") for link in hotel_elements_]
38 Hotel_names = [link.get_attribute("title") for link in hotel_elements_]

```

شکل ۹. استخراج اطلاعات و شاخص‌ها در سلیوم

در مرحله آخر، پس از شناسایی تمام تگ‌ها و کلاس‌ها در سورس کد HTML سایت از حلقه for برای استخراج تمام تگ‌ها و عناصر تمام صفحات استفاده می‌شود و برای ذخیره‌سازی آن در یک فایل اکسل (excel) از کتابخانه pandas در پایتون استفاده می‌شود.

|   | A                       | B           | C         | D          | E               | F              | G         | H              |
|---|-------------------------|-------------|-----------|------------|-----------------|----------------|-----------|----------------|
| 1 | نام هتل                 | امکانات هتل | قیمت اتاق | کیفیت اتاق | موقعیت اقامتگاه | برونکل بهداشتی | رضایتمندی | rate           |
| 2 | هتل پارسیان آزادی تهران | 8/4         | 7/4       | 8/2        | 8/8             | 8/4            | 8/396     | هتل پنج ستاره  |
| 3 | هتل پرشین پلازا تهران   | 7/5         | 6/5       | 7/2        | 7/8             | 7              | 7/378     | هتل پنج ستاره  |
| 4 | هتل نور حیات تهران      | 2/3         | 5         | 7          | 5/7             | 7/7            | 4/52      | هتل پنج ستاره  |
| 5 | هتل استقلال تهران       | 6/4         | 6/1       | 5/2        | 8/3             | 6/7            | 7/436     | هتل پنج ستاره  |
| 6 | هتل تهران رویال تهران   | 8/2         | 7         | 8/4        | 8/8             | 8/8            | 8/35      | هتل چهار ستاره |
| 7 | هتل اسپیناس پلاس تهران  | 7/8         | 6/3       | 7/2        | 8/1             | 7/7            | 7/8       | هتل پنج ستاره  |
| 8 | هتل اسپیناس بلوار تهران | 7/8         | 6/3       | 7/3        | 8/5             | 7/8            | 7/6       | هتل پنج ستاره  |

شکل ۱۰. خوانش داده‌ها

پس از به دست آمدن داده‌های فوق از طریق وب کاوی با پایتون به تجزیه و تحلیل داده‌ها در شاخص‌های مختلف خواهیم پرداخت. در ابتدا، به ارائه آمارهای توصیفی حاصل از داده‌های به دست آمده می‌پردازیم. پس از آن تفاوت میانگین نظرات کاربران برخط استفاده‌کنندگان و همچنین تفاوت معنی‌داری نظرات را برای درجه‌بندی‌های مختلف اقامتگاه‌ها در شاخص‌های مختلف مورد بررسی قرار خواهیم داد. از آنجایی که نظرات در شاخص‌های مختلف از ۱ تا ۱۰ رتبه‌بندی شده‌اند، برای بررسی تفاوت میانگین نظرات عدد بزرگ‌تر از ۵ در نظر گرفته می‌شود.

## یافته‌ها

اطلاعات واکنشی شده از سایت برحسب اقامتگاه و نظرات کاربران و مهمانان اقامتگاه‌ها استخراج شد، مورد پردازش و بررسی اولیه قرار گرفت و در نهایت مورد تجزیه و تحلیل اطلاعاتی قرار گرفت. شاخص‌های استخراج شده بر اساس نظرات

کاربران شامل ۵ شاخص «امکانات هتل»، «قیمت اتاق»، «کیفیت اتاق»، «موقعیت اقامتگاه» و «پروتکل‌های بهداشتی» است که رتبه دهی کاربران در تمام این شاخص‌های از ۱ تا ۱۰ بوده است. یافته‌های تحقیق با نرم‌افزار jamovi ارائه شده است. دو نوع پردازش بر روی داده‌های تحقیق انجام شده است. در ابتدا آمارهای توصیفی در مورد شاخص‌های استخراج شده بر اساس نظرات کاربران مورد بررسی قرار گرفت. در این مرحله فراوانی نظرات؛ میانگین، میانه، انحراف استاندارد، کمترین و بیشترین مقدار نظرات استخراج شد و نمودارهای فراوانی ترسیم گردید. پس از آن تفاوت میانگین نظرات کاربران برخط استفاده‌کنندگان تجزیه و تحلیل شد. بدین منظور از آزمون t-test تک نمونه‌ای استفاده شد. تجزیه و تحلیل دوم، تفاوت معنی‌داری نظرات کاربران در میان گروه‌های درجه‌بندی (ستاره‌ها) مختلف اقامتگاه‌ها بررسی شد و مورد تجزیه و تحلیل قرار گرفت. در این مرحله از آزمون ANOVA یک مرحله‌ای استفاده شد و آماره‌های Welch's و Fisher's مورد تجزیه و تحلیل قرار گرفتند و در پایان نمودار Q-Q plot برای هر یک از شاخص‌ها ترسیم شد.

### تحلیل شاخص‌های نظرات کاربران در سایت اقامت ۲۴

جدول ۱. آمار توصیفی شاخص‌ها

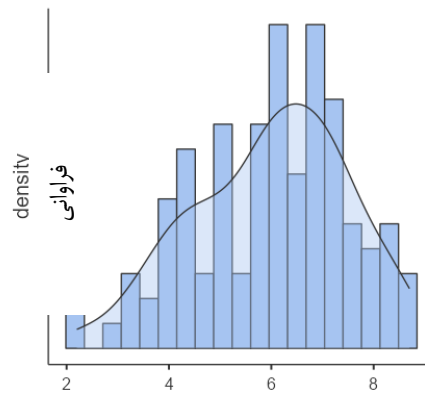
| امکانات هتل | قیمت اتاق | کیفیت اتاق | موقعیت اقامتگاه | پروتکل بهداشتی | فراوانی          |
|-------------|-----------|------------|-----------------|----------------|------------------|
| ۱۰۶         | ۱۰۶       | ۱۰۶        | ۱۰۶             | ۱۰۶            | ۱۰۶              |
| ۵/۹۸        | ۶/۱۱      | ۶/۱۵       | ۷/۴۹            | ۶/۳۸           | میانگین          |
| ۶/۰۵        | ۶/۳۰      | ۶/۲۵       | ۷/۷۰            | ۶/۷۵           | میانه            |
| ۱/۴۹        | ۰/۸۹۴     | ۱/۵۵       | ۱/۱۲            | ۱/۸۳           | انحراف استاندارد |
| ۲/۲۰        | ۱         | ۲/۷۰       | ۳               | ۱              | کمترین           |
| ۸/۷۰        | ۸         | ۹/۳۰       | ۱۰              | ۹/۵۰           | بیشترین          |

جدول (۱)، تجزیه و تحلیل آماری را برای پنج شاخص مربوط به ارزیابی اقامتگاه‌های شهر تهران با توجه به نظرات برخط استفاده‌کنندگان سایت اقامت ۲۴ ارائه می‌دهد: شاخص‌های استخراج شده شامل «امکانات هتل»، «قیمت اتاق»، «کیفیت اتاق»، «موقعیت اقامتگاه» و «پروتکل‌های بهداشتی» است. در اینجا تجزیه و تحلیل دقیق از هر شاخص ارائه می‌شود:

#### امکانات هتل

##### آمار توصیفی و میانگین نظرات

میانگین امتیاز نظرات کاربران برای امکانات رفاهی هتل برابر با ۵/۹۸ است که کمی بالاتر از میانه نظرات (۶/۰۵) است و نشان‌دهنده توزیع نسبتاً متعادل نظرات است. انحراف استاندارد نظرات معادل ۱/۴۹ است که نشان‌دهنده تنوع متوسط در رتبه‌بندی‌ها است. نمرات نظرات از حداقل ۲/۲۰ تا حداکثر ۸/۷۰ متغیر است که طیف وسیعی را در کیفیت درک شده از امکانات اقامتگاه‌ها نشان می‌دهد. شکل (۱۱) نمودار فراوانی نظرات در شاخص «امکانات هتل» را نشان می‌دهد.



طبقه‌بندی امتیازات امکانات هتل

شکل ۱۱. نمودار فراوانی امکانات هتل

جدول (۲) آمار توصیفی امکانات هتل را بر اساس درجه اقامتگاه‌ها (از یک تا پنج ستاره) ارائه می‌دهد. هتل‌های یک ستاره دارای میانگین امتیاز ۵/۸۸ و میانه ۵/۹۰ هستند که نشان‌دهنده توزیع متعادل نظرات در آن‌ها است. هتل‌های دو ستاره کمترین میانگین امتیاز را برای امکانات هتل با ۵/۰۱ و میانه کمی پایین‌تر ۴/۶۰ دارند، که نشان می‌دهد بیش از نیمی از رتبه‌بندی‌ها کمتر از میانگین هستند. هتل‌های چهار ستاره بالاترین میانگین امتیاز را با ۷/۰۱ و میانه ۷/۱۰ دارند که نشان‌دهنده درک بسیار مثبت از امکانات اقامتگاه‌ها است.

جدول ۲. آمار توصیفی امکانات هتل بر اساس درجه اقامتگاه‌ها

| درجه اقامتگاه  | فراوانی | میانگین | میانه | انحراف استاندارد | کمترین | بیشترین |
|----------------|---------|---------|-------|------------------|--------|---------|
| هتل دو ستاره   | ۲۹      | ۵/۰۱    | ۴/۶۰  | ۱/۴۰             | ۲/۲۰   | ۸/۲۰    |
| هتل سه ستاره   | ۳۷      | ۶/۰۷    | ۶/۱۰  | ۱/۲۶             | ۳/۲۰   | ۸/۴۰    |
| هتل پنج ستاره  | ۱۰      | ۶/۵۷    | ۷/۴۵  | ۱/۹۳             | ۲/۳۰   | ۸/۴۰    |
| هتل چهار ستاره | ۱۹      | ۷/۰۱    | ۷/۱۰  | ۱/۰۸             | ۳/۹۰   | ۸/۶۰    |
| هتل یک ستاره   | ۱۱      | ۵/۸۸    | ۵/۹۰  | ۱/۲۲             | ۳/۹۰   | ۸/۷۰    |

در این قسمت تفاوت میانگین نظرات کاربران را با فرض اینکه میانگین جامعه بزرگ‌تر از ۵ است مورد تفسیر قرار می‌دهیم به طوری که در فرضیه صفر ( $H_0$ ) میانگین نظرات برابر با ۵ است و در فرضیه یک ( $H_1$ ) میانگین نظرات بزرگ‌تر از ۵ است. بر اساس اطلاعات به دست آمده از جدول (۳) مقدار  $p$  کمتر از ۰/۰۰۱ است که کمتر از سطح معنی‌داری رایج یعنی (۰/۰۵) است. این بدان معناست که فرضیه صفر با درجه بالایی از اطمینان رد می‌شود و مقدار  $t$ -value برابر با ۶/۷۳ است که نشان می‌دهد میانگین نظرات به طور قابل توجهی از میانگین فرض (۵) بیشتر است. هم‌چنین تفاوت میانگین نشان می‌دهد که امکانات هتل برابر با ۰/۹۷۵ از میانگین فرضیه بالاتر است.

جدول ۳. T-Test تک نمونه‌ای امکانات هتل

| آماره       | درجه آزادی | مقدار P | اختلاف میانگین |
|-------------|------------|---------|----------------|
| Student's t | ۶/۷۳       | ۰/۰۰۱ < | ۰/۹۷۵          |

Note.  $H_a \mu > 5$

تحلیل معناداری نظرات برخط کاربران برای شاخص امکانات هتل در میان درجه بندی اقامتگاه‌ها هر دو آزمون ANOVA (ولش و فیشر) نتایج معنی داری را نشان می‌دهند ( $p < 0/001$ )، به این معنی که فرضیه صفر را در هر دو مورد رد می‌کنیم. این به شدت نشان می‌دهد که حداقل یک گروه دارای میانگین امتیاز متفاوتی برای امکانات هتل در مقایسه با سایرین است. نتایج قابل توجه نشان می‌دهد که رتبه بندی برای امکانات هتل به طور قابل توجهی در گروه‌های مختلف مورد مقایسه متفاوت است. از این رو، نتایج ANOVA یک طرفه بر اساس جدول (۴) نشان می‌دهد که تفاوت‌های قابل توجهی در رتبه بندی امکانات هتل در بین گروه‌های مختلف وجود دارد.

جدول ۴. آزمون ANOVA یک طرفه امکانات هتل برای درجه‌های اقامتگاه‌ها

| ANOVA - یک طرفه                                       |          |       |     |       |         |
|---|----------|-------|-----|-------|---------|
|   |          | F     | df1 | df2   | p       |
| امکانات هتل   | Welch's  | ۷/۴۹  | ۴   | ۳۲/۸  | ۰/۰۰۱ < |
|   | Fisher's | ۷/۱۱  | ۴   | ۱۰۱   | ۰/۰۰۱ < |
| آزمون نرمال بودن (Shapiro-Wilk)                       |          |       |     |       |         |
| امکانات هتل   | W        | p     |     |       |         |
|   | ۰/۹۸۲    | ۰/۱۴۷ |     |       |         |
| نکته. مقدار p پایین نشان دهنده نقض فرض نرمال بودن است |          |       |     |       |         |
| آزمون همگنی واریانس‌ها (Levene's)                     |          |       |     |       |         |
| امکانات هتل   | F        | df1   | df2 | p     |         |
|   | ۱/۱۴     | ۴     | ۱۰۱ | ۰/۳۴۴ |         |

### قیمت اتاق

#### آمار توصیفی و میانگین نظرات

نظرات کاربران در مورد قیمت اتاق دارای میانگین امتیاز ۶/۱۱ و میانه ۶/۳۰ است که نشان دهنده توزیع کمی مایل به سمت امتیازات پایین تر است. انحراف معیار نسبتاً پایین که برابر با ۰/۸۹۴ است نشان می‌دهد اکثر رتبه بندی‌ها نزدیک به میانگین هستند. رتبه بندی‌ها از حداقل ۱ تا بالاترین ۸ متغیر است، که نشان می‌دهد که قیمت اتاق‌ها به طور کلی معقول و با موارد دور افتاده اندک تعلق می‌شوند.

جدول ۵. آمار توصیفی قیمت اتاق بر اساس درجه اقامتگاه‌ها

| بیشترین | کمترین | انحراف استاندارد | میانه | میانگین | فراوانی | درجه اقامتگاه  |
|---------|--------|------------------|-------|---------|---------|----------------|
| ۷/۲۰    | ۱      | ۱/۲۰۸            | ۵/۸۰  | ۵/۶۳    | ۲۹      | هتل دو ستاره   |
| ۸       | ۵      | ۰/۶۵۷            | ۶/۳۰  | ۶/۲۴    | ۳۷      | هتل سه ستاره   |
| ۷/۴۰    | ۵      | ۰/۷۸۰            | ۶/۴۰  | ۶/۳۷    | ۱۰      | هتل پنج ستاره  |
| ۷/۱۰    | ۵/۳۰   | ۰/۴۴۳            | ۶/۶۰  | ۶/۵۵    | ۱۹      | هتل چهار ستاره |
| ۷/۱۰    | ۴/۷۰   | ۰/۷۹۱            | ۶     | ۵/۹۱    | ۱۱      | هتل یک ستاره   |

بر اساس جدول (۵) هتل‌های یک ستاره دارای میانگین قیمت اتاق ۵/۹۱ و میانه ۶ هستند که نشان دهنده توزیع متعادل نظرات است. میانگین امتیاز هتل‌های دو ستاره در خصوص قیمت اتاق برابر با ۵/۶۳ و با میانه کمی بالاتر از میانگین برابر با ۵/۸۰ است که نشان می‌دهد بیش از نیمی از رتبه بندی‌ها بالاتر از میانگین هستند. هتل‌های سه ستاره با میانگین اتاق ۶/۳۷ و میانه ۶/۳۰ هستند که نشان دهنده توزیع متعادل در اطراف میانگین است. هتل‌های چهار ستاره بالاترین میانگین

قیمت اتاق را با ۶/۵۵ و میانه ۶/۶۰ دارند که این مسئله نشان دهنده درک بسیار مثبت از قیمت اتاق‌ها است. هتل‌های پنج ستاره دارای میانگین ۶/۳۷ و میانه ۶/۴۰ هستند که نشان می‌دهد توزیع نظرات نسبتاً متقارن است. در این قسمت تفاوت میانگین نظرات کاربران را با فرض اینکه میانگین جامعه بزرگ‌تر از ۵ است مورد تفسیر قرار می‌دهیم به طوری که در فرضیه صفر ( $H_0$ ) میانگین نظرات برابر با ۵ است و در فرضیه یک ( $H_1$ ) میانگین نظرات بزرگ‌تر از ۵ است.

بر اساس اطلاعات به دست آمده از جدول (۶) مقدار  $p$  کمتر از ۰,۰۰۱ است که کمتر از سطح معنی‌داری رایج یعنی (۰/۰۵) است. این بدان معناست که فرضیه صفر با درجه بالایی از اطمینان رد می‌شود و مقدار  $t$ -value برابر با ۱۲/۸ است که نشان می‌دهد میانگین نظرات به طور قابل توجهی از میانگین فرض (۵) بیشتر است. همچنین تفاوت میانگین نشان می‌دهد که قیمت اتاق برابر با ۱/۱۱ به مقدار متوسطی از میانگین فرضیه بالاتر است. این نشان دهنده انحراف قابل توجهی در قیمت اتاق‌ها از میانگین فرضیه است. جدول (۶) مقادیر سطح متوسط میانگین نظرات را نشان می‌دهد.

جدول ۶. T-Test تک نمونه‌ای قیمت اتاق

| آماره       | درجه آزادی | مقدار P | اختلاف میانگین |
|-------------|------------|---------|----------------|
| Student's t | ۱۰۵        | < ۰/۰۰۱ | ۱/۱۱           |

Note.  $H_a \mu > 5$

تحلیل معناداری نظرات برخط کاربران برای شاخص قیمت اتاق در میان درجه‌بندی اقامتگاه‌ها از آنجایی که هر دو آزمون ANOVA ولش و فیشر دارای مقادیر  $p$  کمتر از ۰/۰۵ (به ترتیب ۰/۰۰۷ و ۰/۰۰۳) هستند، شواهد کافی برای رد فرضیه صفر میانگین گروهی برابر داریم. این نشان می‌دهد که تفاوت قابل توجهی در قیمت اتاق‌ها در بین گروه‌ها وجود دارد. بر اساس جدول (۷)، هر دو آزمون ANOVA (ولش و فیشر) نتایج معنی‌داری را نشان می‌دهند ( $p < 0/05$ )، به این معنی که در هر دو مورد فرضیه صفر را رد می‌کنیم.

جدول ۷. آزمون ANOVA یک‌طرفه قیمت اتاق برای درجه‌های اقامتگاه‌ها

| ANOVA - یک‌طرفه   |          |      |     |         |         |
|---|----------|------|-----|---------|---------|
|   | F        | df1  | df2 | p       |         |
| قیمت اتاق   | Welch's  | ۴/۲۹ | ۴   | ۳۳      | < ۰/۰۰۷ |
|   | Fisher's | ۴/۳۲ | ۴   | ۱۰۱     | < ۰/۰۰۳ |
| آزمون نرمال بودن (Shapiro-Wilk)                         |          |      |     |         |         |
| قیمت اتاق   | W        |      |     | p       |         |
|   | ۰/۸۹۴    |      |     | < ۰/۰۰۱ |         |
| نکته. مقدار $p$ پایین نشان دهنده نقض فرض نرمال بودن است |          |      |     |         |         |
| آزمون همگنی واریانس‌ها (Levene's)                       |          |      |     |         |         |
|   | F        | df1  | df2 | p       |         |
| قیمت اتاق   | ۲/۲۶     | ۴    | ۱۰۱ | ۰/۰۶۷   |         |

تحلیل معناداری نظرات برخط کاربران برای شاخص کیفیت اتاق در میان درجه‌بندی اقامتگاه‌ها هر دو آزمون ANOVA (ولش و فیشر) نتایج بسیار معنی‌داری را نشان می‌دهند ( $p < 0/001$ )، که نشان می‌دهد ما فرضیه صفر را در هر دو مورد رد می‌کنیم. این نشان می‌دهد که حداقل یک گروه در مقایسه با گروه‌های دیگر دارای میانگین رتبه‌بندی کیفیت اتاق متفاوت است. نتایج قابل توجه نشان می‌دهد که رتبه‌بندی کیفیت اتاق به طور قابل توجهی



در گروه‌های مختلف مقایسه شده متفاوت است. نتایج ANOVA یک‌طرفه برای رتبه‌بندی کیفیت اتاق مطابق با جدول (۱۰) تفاوت‌های قابل توجهی را در بین گروه‌های مختلف نشان می‌دهد.

جدول ۱۰. آزمون ANOVA یک‌طرفه کیفیت اتاق برای درجه‌های اقامتگاه‌ها

| ANOVA - یک‌طرفه                                       |          |      |     |       |         |
|---|----------|------|-----|-------|---------|
|   | F        | df1  | df2 | p     |         |
| قیمت اتاق   | Welch's  | ۶/۶۵ | ۴   | ۳۳/۸  | < ۰/۰۰۱ |
|   | Fisher's | ۶/۰۷ | ۴   | ۱۰۱   | < ۰/۰۰۱ |
| آزمون نرمال بودن (Shapiro-Wilk)                       |          |      |     |       |         |
| قیمت اتاق   | W        |      |     | p     |         |
|   | ۰/۸۹۴    |      |     | ۰/۶۶۹ |         |
| نکته. مقدار p پایین نشان‌دهنده نقض فرض نرمال بودن است |          |      |     |       |         |
| آزمون همگنی واریانس‌ها (Levene's)                     |          |      |     |       |         |
| قیمت اتاق   | F        | df1  | df2 | p     |         |
|   | ۱/۰۸     | ۴    | ۱۰۱ | ۰/۳۷۲ |         |

### موقعیت اقامتگاه‌ها

#### آمار توصیفی و میانگین نظرات

مکان دارای بالاترین میانگین امتیاز با ۷/۴۹ و میانه ۷/۷۰ است که نشان می‌دهد اکثر هتل‌ها در موقعیت مطلوبی قرار دارند. انحراف استاندارد شاخص برابر با ۱/۱۲ است که نشان می‌دهد رتبه‌بندی‌ها به‌طور کلی نزدیک به میانگین هستند. امتیازات از ۳ تا ۱۰ کامل متغیر است، که نشان می‌دهد درحالی‌که بیشتر مکان‌ها رتبه‌بندی بالایی دارند، اما چند مکان کمتر مطلوب نیز وجود دارد.

بر اساس جدول (۱۱)، میانگین رتبه‌بندی هتل‌های یک ستاره ۷/۰۱ و میانه آن‌ها برابر با ۷ است که نشان‌دهنده توزیع متعادل است. انحراف استاندارد برابر با ۱/۵۰۸ که بالاترین میزان را در بین تمام طبقه‌ها دارد و تنوع قابل توجهی را نشان می‌دهد. میانگین رتبه‌بندی هتل‌های دو ستاره برابر با ۶/۹۲ و بسیار نزدیک به میانه برابر با ۶/۹۰ است که نشان‌دهنده توزیع متعادل در اطراف میانگین است. هتل‌های سه ستاره دارای رتبه میانگین برابر با ۷/۶۱ و میانه ۷/۶۰ هستند که نشان‌دهنده توزیع بسیار متعادل است. هتل‌های چهار ستاره بالاترین میانگین رتبه‌بندی موقعیت مکانی را با ۸/۲۱ و میانه ۸/۱۰ دارند که نشان‌دهنده درک بسیار مثبت از موقعیت مکانی است. میانگین رتبه‌بندی هتل‌های پنج ستاره ۷/۹۳ و میانه بالاتر از میانگین برابر با ۸/۳۰ است که نشان می‌دهد بیش از نیمی از رتبه‌بندی‌ها بالاتر از میانگین هستند و همچنین نشان‌دهنده یک انحراف مثبت است. انحراف استاندارد برابر با ۰/۹۴۴ است که تنوع متوسطی را نشان می‌دهد.

جدول ۱۱. آمار توصیفی موقعیت اقامتگاه بر اساس درجه اقامتگاه‌ها

| درجه اقامتگاه   | فراوانی | میانگین | میانه | انحراف استاندارد | کمترین | بیشترین |
|-----------------|---------|---------|-------|------------------|--------|---------|
| هتل دو ستاره    | ۲۹      | ۶/۹۲    | ۶/۹۰  | ۱/۲۳۳            | ۳      | ۹       |
| هتل سه ستاره    | ۳۷      | ۷/۶۱    | ۷/۶۰  | ۰/۹۱۰            | ۴/۴۰   | ۹       |
| موقعیت اقامتگاه | ۱۰      | ۷/۹۳    | ۸/۳۰  | ۰/۹۴۴            | ۵/۷۰   | ۸/۸۰    |
| هتل چهار ستاره  | ۱۹      | ۸/۲۱    | ۸/۱۰  | ۰/۴۸۰            | ۶/۸۰   | ۸/۹۰    |
| هتل یک ستاره    | ۱۱      | ۷/۰۱    | ۷     | ۱/۵۰۸            | ۴/۸۰   | ۱۰      |

آزمون t تک نمونه‌ای انجام شده بر روی رتبه‌بندی موقعیت مکان‌های اقامتی مطابق جدول (۱۲)، انحراف معنی‌داری را از میانگین فرضی ۵ در جهت مثبت نشان می‌دهد. مقدار t-value برابر با ۲۲/۹ و مقدار p-value کمتر از ۰/۰۰۱ است که فرضیه صفر رد می‌شود و نتیجه می‌گیریم که میانگین رتبه برای موقعیت مکانی اقامتگاه‌ها به‌طور قابل توجهی بیشتر از ۵ است. اختلاف میانگین برابر با ۲/۴۹ است که این تفاوت را کمی بیشتر می‌کند و نشان‌دهنده این است که مکان اقامتگاه به‌طور قابل توجهی از ارزش فرضی ۵ بالاتر است.

جدول ۱۲. T-Test تک نمونه‌ای موقعیت اقامتگاه

| آماره       | درجه آزادی | مقدار P | اختلاف میانگین |
|-------------|------------|---------|----------------|
| Student's t | ۲۲/۹       | < ۰/۰۰۱ | ۲/۴۹           |

Note.  $H_a \mu > 5$

### تحلیل معناداری نظرات برخط کاربران برای شاخص موقعیت اقامتگاه در میان درجه‌بندی اقامتگاه‌ها

بر اساس جدول (۱۳)، نتایج ANOVA یک‌طرفه برای رتبه‌بندی موقعیت اقامتگاه تفاوت‌های قابل توجهی را در بین گروه‌های مختلف نشان می‌دهد. هر دو آنالیز واریانس Welch و Fisher's ANOVA (به ترتیب ۷/۷۹ و ۵/۷۰) با مقادیر p کمتر از ۰/۰۰۱ تولید کردند که ما را به رد فرضیه صفر میانگین گروهی مساوی سوق داد. این نشان می‌دهد که حداقل یک گروه دارای میانگین رتبه‌بندی محل اقامت متفاوت است.

از آنجایی که هر دو آزمون ANOVA ولش و فیشر دارای p-value کمتر از ۰/۰۰۱ هستند، ما شواهد قوی برای رد فرضیه صفر میانگین گروهی برابر داریم. این نشان می‌دهد که تفاوت‌های قابل توجهی در رتبه‌بندی محل اقامت در بین گروه‌ها وجود دارد.

جدول ۱۳. آزمون ANOVA یک‌طرفه موقعیت اقامتگاه برای درجه‌های اقامتگاه‌ها

| ANOVA - یک‌طرفه                                       |          |      |     |         |         |
|---|----------|------|-----|---------|---------|
|   | F        | df1  | df2 | p       |         |
| قیمت اتاق   | Welch's  | ۷/۷۹ | ۴   | ۳۲/۵    | < ۰/۰۰۱ |
|   | Fisher's | ۵,۷۰ | ۴   | ۱۰۱     | < ۰/۰۰۱ |
| آزمون نرمال بودن (Shapiro-Wilk)                       |          |      |     |         |         |
| قیمت اتاق   | W        |      |     | p       |         |
|   | ۰/۹۴۹    |      |     | < ۰/۰۰۱ |         |
| نکته. مقدار p پایین نشان‌دهنده نقض فرض نرمال بودن است |          |      |     |         |         |
| آزمون همگنی واریانس‌ها (Levene's)                     |          |      |     |         |         |
| قیمت اتاق   | F        | df1  | df2 | p       |         |
|   | ۲/۹۸     | ۴    | ۱۰۱ | ۰/۰۲۳   |         |

این مطالعه با ارائه یک روش جامع و هوشمند برای وب کاوی نظرات برخط کاربران درباره اقامتگاه‌های شهر تهران، به نیاز مهمی در صنعت گردشگری پاسخ می‌دهد. نتایج به‌دست‌آمده از این تحقیق نشان می‌دهد که استفاده از تکنیک‌های وب کاوی می‌تواند به‌طور مؤثری داده‌های بزرگ و پیچیده را تجزیه و تحلیل کند و از این طریق به ارائه بینش‌های ارزشمندی برای تصمیم‌گیری‌های استراتژیک در زمینه مدیریت اقامتگاه‌ها و خدمات گردشگری منجر شود. این امر در مطالعات نویسندگان گذشته از جمله Han & Anderson (2021)، Choong (2019)، Barbera و همکارانش (2023) و Adhinugroho و همکارانش (2020) تأکید شده است.

از مهم‌ترین دستاوردهای این پژوهش می‌توان به امکان دستیابی به تحلیل دقیق‌تری از نظرات کاربران و درک بهتر از عوامل مؤثر بر رضایت و نارضایتی آن‌ها اشاره کرد. نتایج نشان داد که نظرات مثبت کاربران عمدتاً بر روی عواملی نظیر موقعیت مکانی، امکانات و خدمات اقامتگاه‌ها متمرکز بوده، درحالی‌که نارضایتی‌ها بیشتر به موضوعات ملموس مانند قیمت و کیفیت فیزیکی اتاق‌ها مرتبط بوده است.

این تحقیق همچنین نشان می‌دهد که تکنیک وب کاوی می‌تواند به‌عنوان یک ابزار کارآمد برای مدیریت و بازاریابی اقامتگاه‌ها استفاده شود. با تحلیل نظرات کاربران، مدیران اقامتگاه‌ها می‌توانند به اطلاعات دقیق و به‌روز درباره نیازها و ترجیحات مشتریان دست یابند و بر این اساس خدمات خود را بهبود دهند. این پژوهش با موفقیت به یکی از چالش‌های مهم در حوزه گردشگری و مدیریت اقامتگاه‌ها پاسخ می‌دهد؛ یعنی تحلیل و درک عمیق‌تر از نظرات کاربران در فضای آنلاین.

نتایج این تحقیق نه تنها اهمیت ویژه‌ای در سطح مدیریتی دارد، بلکه از نظر نظری نیز به غنای ادبیات موجود در زمینه گردشگری دیجیتال و تحلیل داده‌های اجتماعی می‌افزاید. شواهد به‌دست‌آمده نشان می‌دهد که نظرات کاربران می‌توانند به‌عنوان منابع غنی از دانش مورد استفاده قرار گیرند که در صورت پردازش صحیح، می‌توانند به مدیران اقامتگاه‌ها بینش‌های ارزشمندی ارائه دهند. این بینش‌ها شامل شناسایی نقاط قوت و ضعف در خدمات، تعیین اولویت‌های مشتریان، و نیز درک عمیق‌تری از انتظارات آن‌ها می‌شود.

در نهایت، این پژوهش با توجه به نوآوری در استفاده از وب کاوی در زمینه تحلیل نظرات کاربران اقامتگاه‌های تهران، پایه‌های محکمی برای تحقیقات آتی در این حوزه فراهم می‌کند. به‌خصوص در زمینه‌هایی نظیر بهبود تجربه مشتری، بهینه‌سازی قیمت‌گذاری و ارتقاء کیفیت خدمات اقامتگاه‌ها، می‌توان از یافته‌های این پژوهش به‌عنوان مبنای تصمیم‌گیری‌های استراتژیک استفاده کرد. این موارد در تحقیق‌های سایر نویسندگان نظیر Fazzolari (2018)، Adhinugroho و همکاران (2020)، Nwakanma و همکاران (2019) مشاهده می‌شود.

تحقیق حاضر همچنین از منظر عملی، پیامدهای چشمگیری برای صنعت گردشگری دارد. با استفاده از نتایج این تحقیق، مدیران می‌توانند استراتژی‌های بازاریابی خود را بهبود بخشند، قیمت‌گذاری‌ها را بهینه کنند و خدمات را به‌گونه‌ای طراحی کنند که بیشترین رضایت را در مشتریان ایجاد کند. به‌علاوه، این تحقیق نشان می‌دهد که تحلیل نظرات کاربران می‌تواند به‌طور مستقیم به افزایش رقابت‌پذیری اقامتگاه‌ها منجر شود؛ چرا که مدیریت مؤثر نظرات و انتقادات آنلاین به شهرت برند و جذب مشتریان بیشتر کمک شایانی می‌کند.

در سطح کلان‌تر، این پژوهش می‌تواند به‌عنوان یک چارچوب پایه برای مطالعات آتی در حوزه‌های مختلف از جمله هوش مصنوعی، تحلیل احساسات، و مدیریت تجربه مشتری به کار گرفته شود. با توجه به اینکه صنعت گردشگری به‌شدت تحت تأثیر نظرات و بازخوردهای آنلاین است، این تحقیق نشان می‌دهد که چگونه می‌توان از داده‌های تولیدشده توسط کاربران به‌عنوان یک منبع استراتژیک بهره برد و به‌این‌ترتیب، ارزش‌افزوده قابل‌توجهی برای سازمان‌ها ایجاد کرد.

در نهایت، این مطالعه با نشان دادن قدرت و قابلیت‌های وب کاوی در درک بهتر از دینامیک‌های مشتریان، راه را برای نوآوری‌های بیشتر در این حوزه هموار می‌کند. آینده‌پژوهی و تحقیقات بیشتر در این زمینه می‌تواند به بهبود مستمر خدمات و تجربه مشتریان منجر شود و بدین ترتیب، نقشی کلیدی در موفقیت و پیشرفت صنعت گردشگری ایفا کند.

## نتیجه گیری

تأثیر بررسی‌ها و نظرات آنلاین کاربران و گردشگران در عصر دیجیتال امروز، هم برای کسب‌وکارها و هم برای مصرف‌کنندگان، روزبه‌روز اهمیت بیشتری پیدا می‌کند. گردشگران برای تصمیم‌گیری در مورد برنامه‌ریزی سفر به نظرات آنلاین اعتماد می‌کنند، درحالی‌که کسب‌وکارها از آن‌ها برای ایجاد استراتژی‌های بازاریابی مؤثر استفاده می‌کنند. با این وجود، حجم زیاد داده‌های موجود، تجزیه و تحلیل تک‌تک نظرات را غیرعملی می‌کند. بنابراین، در سال‌های اخیر، تلاش‌های زیادی برای ارائه روش‌هایی برای تجزیه و تحلیل خودکار و خلاصه‌سازی نظرات انجام شده است.

پژوهش‌های گذشته بر استفاده از تکنیک‌های خودکار برای کسب دانش جدید و مفید از داده‌های مرتبط با گردشگری تأکید کرده‌اند. بنابراین، این پژوهش همسو با پژوهش‌های گذشته است و نشان می‌دهد که چگونه تحلیل‌های مختلف انجام شده بر روی داده‌های مرتبط با نظرات کاربران از سایت‌های آنلاین در مورد اقامتگاه‌های شهر تهران، به کسب‌وکارها و کاربران کمک می‌کند و منجر به بهبود فضای کسب‌وکار می‌شود.

با مطالعه و تحلیل این نظرات، می‌توان نیازها و ترجیحات گردشگران را بهتر درک کرد و بهبودهای لازم در اقامتگاه‌ها را انجام داد. در زیر به برخی از تأثیرات این تحلیل‌ها از روش‌های خودکار وب کاوی برای کسب‌وکارها و کاربران می‌توان اشاره کرد:

۱. **بهبود کیفیت خدمات:** تحلیل نظرات آنلاین کاربران به کسب‌وکارها کمک می‌کند تا نقاط ضعف خود را شناسایی کنند و در بهبود آن اقداماتی را انجام دهند.

۲. **تصمیم‌گیری بهتر کاربران:** نظرات آنلاین به گردشگران کمک می‌کند تا قبل از سفر به تهران، اطلاعات کامل‌تری درباره اقامتگاه‌ها و تجربه‌های دیگران دریافت کنند. این اطلاعات به آن‌ها کمک می‌کند تا تصمیم بهتری در مورد انتخاب اقامتگاه بگیرند و تجربه مثبت‌تری داشته باشند.

۳. **استراتژی‌های بازاریابی:** تحلیل نظرات کاربران به کسب‌وکارها اطلاعات قابل‌اعتمادی را درباره نقاط قوت و ضعف خود و رقبا فراهم می‌کند. این اطلاعات می‌تواند به کسب‌وکارها در تدوین استراتژی‌های بازاریابی مؤثر کمک کند.

۴. **ارتقای اعتماد به کسب‌وکار:** نظرات آنلاین مثبت کاربران و گردشگران درباره اقامتگاه‌ها و خدمات ارائه شده، به کسب‌وکارها کمک می‌کند تا اعتماد مشتریان را به دست آورند. اعتماد به کسب‌وکار می‌تواند منجر به افزایش تکراری و معرفی مشتریان جدید شود.

بر این اساس، در مطالعه پیش رو، به منظور جمع‌آوری داده‌های نظرات کاربران از اقامتگاه‌های شهر تهران مراحل بدین شرح انجام پذیرفته است: ۱- بررسی سایت‌ها و انتخاب سایت مورد نظر؛ ۲- بررسی ساختار سایت و انتخاب تگ‌هایی که قرار است استخراج شود؛ ۳- بررسی عناصر و تگ‌های HTML و CSS عناصر سایت؛ ۴- انتخاب ابزارهای فنی مورد نیاز با توجه به عناصر؛ ۵- کد نویسی استخراج داده‌ها؛ ۶- تحلیل نظرات آنلاین مستخرج

از این رو، این پژوهش با رویکرد نوآورانه خود، می‌تواند به بحث درباره پتانسیل‌های وب کاوی برای تحلیل شناختی داده‌های حجیم در بخش گردشگری کمک کند. ابزارهای خودکار وب کاوی می‌توانند به روشی بسیار سریع و شهودی به جستجوی داده‌های وب در این بخش بپردازند بدون آنکه با مشکلات قانونی مانند عدم رضایت برای دسترسی به اطلاعات خصوصی مواجه شوند. این ابزارها می‌توانند، به مشاوره در صنعت گردشگری نیز کمک کنند، زیرا می‌توانند استخراج

داده‌های رقبا و سایر اطلاعات مرتبط را بهینه کنند این دانش و پایش لحظه‌ای آن می‌تواند به کسب‌وکارهای گردشگری مزیت رقابتی بدهد. به طوری که آن‌ها به سرعت و به صراحت، داده‌هایی در مورد عملکرد هر کسب‌وکار ارائه می‌دهند.

### حامی مالی

این اثر حامی مالی نداشته است.

### سهام نویسندگان در پژوهش

نویسندگان در تمام مراحل و بخش‌های انجام پژوهش سهم برابر داشتند.

### تضاد منافع

نویسندگان اعلام می‌دارند که هیچ تضاد منافی در رابطه با نویسندگی و یا انتشار این مقاله ندارند.

### تقدیر و تشکر

نویسندگان از همه کسانی که در انجام این پژوهش به ما یاری رساندند، به ویژه کسانی که کار ارزیابی کیفیت مقالات را انجام دادند، تشکر و قدردانی می‌نمایند.

### منابع

- حق وردی زاده، ابوالفضل؛ زارعی، قاسم؛ عسگر نژاد نوری، باقر و رحیمی کلور، حسین. (۱۴۰۲). الگوی توسعه گردشگری هوشمند شهری مبتنی بر اکوسیستم بازاریابی مطالعه موردی: شهر تبریز. *مجله گردشگری شهری*، ۱۰(۳)، ۱۲۷-۱۴۶. DOI: 10.22059/jut.2023.365511.1157
- میمندی، فاطمه؛ کیوان نیا، سامان و شیرمحمدی، علی، (۱۴۰۲). بررسی چالش‌های هوشمند سازی اقامتگاه‌های بوم گردی. *مجله گردشگری شهری*، ۱۰(۴)، ۱۷-۳۵. DOI: 10.22059/jut.2024.352721.1102

### References

- Adhinugroho, Y., Putra, A. P., Luqman, M., Ermawan, G. Y., Mariyah, S., & Pramana, S. (2020). Development of online travel Web scraping for tourism statistics in Indonesia. DOI: 10.47989/irpaper885
- Ahmad, W., & Sun, J. (2018). Modeling consumer distrust of online hotel reviews. *International Journal of Hospitality Management*, 71, 77-90. DOI: 10.1016/j.ijhm.2017.12.005
- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of travel research*, 58(2), 175-191. DOI: 10.1177/004728751774775
- Ali, T., Omar, B., & Soulaïmane, K. (2022). Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *MethodsX*, 9, 101894. DOI: 10.1016/j.mex.2022.101894
- Aluri, A., Price, B. S., & McIntyre, N. H. (2019). Using machine learning to cocreate value through dynamic customer engagement in a brand loyalty program. *Journal of Hospitality & Tourism Research*, 43(1), 78-100. DOI: 10.1177/1096348017753521
- Barbera, G., Araujo, L., & Fernandes, S. (2023). The Value of Web Data Scraping: An Application to TripAdvisor. *Big Data and Cognitive Computing*, 7(3), 121. DOI: 10.3390/bdcc7030121
- Beck, J., Danilenko, M., Sperber, L., Wiersma, B., & Egger, R. (2017). Connecting big data and service quality evaluation—developing a service quality map of the Austrian hospitality industry through the application of big data. *The Gaze: Journal of Tourism and Hospitality*, 8, 40-54. DOI: 10.3126/gaze.v8i0.17831
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1-24. DOI: 10.1080/19368623.2015.983631

- Choong, W. J. (2019). *An automated web scraping tool for Malaysia tourism* (Doctoral dissertation, UTAR).
- Dogucu, M., & Çetinkaya-Rundel, M. (2021). Web scraping in the statistics and data science curriculum: Challenges and opportunities. *Journal of Statistics and Data Science Education*, 29(sup1), S112-S122. DOI: 10.1080/10691898.2020.1787116  
DOI: 10.1080/13683500.2015.1127336
- Fazzolari, M., & Petrocchi, M. (2018). A study on online travel reviews through intelligent data analysis. *Information Technology & Tourism*, 20(1), 37-58. DOI: 10.1007/s40558-018-0121-z
- Gheorghe, M., Mihai, F. C., & Dârdală, M. (2018). Modern techniques of web scraping for data scientists. *International Journal of User-System Interaction*, 11(1), 63-75.
- Haghverdizadeh, A., Zarei, G., Asgarnezhad Nouri, B., & Rahimi Kolour, H. (2023). Urban Smart Tourism Development Model Based on Marketing Ecosystem the Case Study of Tabriz City. *Journal of urban tourism*, 10(3), 127-146. DOI: 10.22059/jut.2023.365511.1157 [In Persian]
- Han, S., & Anderson, C. K. (2021). Web scraping for hospitality research: Overview, opportunities, and implications. *Cornell Hospitality Quarterly*, 62(1), 89-104. DOI: 10.1177/193896552097358
- Kamarazaman, N., Ali, N., & Arshad, H. (2024). Leveraging Web Scraping To Gather Tourism Information Data. *Journal of Event, Tourism and Hospitality Studies*, 4, 16-29. DOI: 10.32890/jeth2024.4.2
- Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3). DOI: 10.32890/jeth2024.4.2
- Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 59, 554-563. DOI: 10.1016/j.tourman.2016.08.012
- Luo, N., Kwan, C., Sun, Y., & Zhang, F. (2020, October). Analyzing and Filtering Food Items In Restaurant Reviews: Sentiment Analysis and Web Scraping. In *Computer Science & Information Technology (CS & IT) Computer Science Conference* DOI: 10.5121/csit.2020.101208
- Meymandi, F., Keyvannia, S., & Shirmohammadi, A. (2024). Examining the Challenges of Smartening Eco-lodges. *Journal of urban tourism*, 10(4), 17-35. DOI: 10.22059/jut.2024.352721.1102 [In Persian]
- Moro, S., Batista, F., Rita, P., Oliveira, C., & Ribeiro, R. (2019a). Are the states united? An analysis of US hotels' offers through TripAdvisor's eyes. *Journal of Hospitality & Tourism Research*, 43(7), 1112-1129. DOI: 10.1177/109634801985479
- Moro, S., Ramos, P., Esmerado, J., & Jalali, S. M. J. (2019b). Can we trace back hotel online reviews' characteristics using gamification features? *International Journal of Information Management*, 44, 88-95. DOI: 10.1016/j.ijinfomgt.2018.09.015
- Moro, S., Rita, P., & Oliveira, C. (2018). Factors influencing hotels' online prices. *Journal of Hospitality Marketing & Management*, 27(4), 443-464. DOI: 10.1080/19368623.2018.1395379
- Nwakanma, C. I., Ogbonna, A. C., Etus, C., Nwifor, E. U., Onyebuchi, J. E., & Ugwueke, E. C. Predictive analytics of customer sentiments towards Nigerian hospitality industry: Case study approach. In *Proc. 3<sup>rd</sup> International Conference on Intelligent Computing and Emerging Technologies (ICET 2019)* (pp. 60-68).
- Oses Fernández, N., Kepa Gerrikagoitia, J., & Alzua-Sorzabal, A. (2018). Sampling method for monitoring the alternative accommodation market. *Current Issues in Tourism*, 21(7), 721-734.
- Pang, C. C. (2023). *Hotel recommendation system with machine learning* (Doctoral dissertation, UTAR).
- Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the impact of online reviews on hotel performance: an empirical analysis. *Journal of travel research*, 56(2), 235-249. DOI: 10.1177/0047287516636
- Pokhrel, S., Somasiri, N., Jeyavadhanam, R., & Ganesan, S. (2023). Web Data Scraping Technology Using Term Frequency Inverse Document Frequency to Enhance the Big Data Quality on Sentiment Analysis. *International Journal of Electrical and Computer Engineering*, 17(11), 300-307.

- Radojevic, T., Stanic, N., & Stanic, N. (2015). Ensuring positive feedback: Factors that influence customer satisfaction in the contemporary hospitality industry. *Tourism management*, 51, 13-21. DOI: [10.1016/j.tourman.2015.04.002](https://doi.org/10.1016/j.tourman.2015.04.002)
- Rajan, A., & Shyam, A. V. (2015). Sentiment Analysis on Customer Reviews in Tourism-A Text Mining Approach.
- Renganathan, V., & Upadhya, A. (2021). Dubai restaurants: A sentiment analysis of tourist reviews. *Academica Turistica-Tourism and Innovation Journal*, 14(2). DOI: [10.26493/2335-4194.14.165-174](https://doi.org/10.26493/2335-4194.14.165-174)
- Sahu, S., Divya, K., Rastogi, N., Yadav, P. K., & Perwej, Y. (2022). Sentimental Analysis on Web Scraping Using Machine Learning Method. *Journal of Information and Computational Science (JOICS)*, ISSN, 1548-7741. DOI: [10.12733/JOICS.2022/V12I08.535569.67004](https://doi.org/10.12733/JOICS.2022/V12I08.535569.67004)
- Sequeira, S., Joy, J., Dsouza, D., & Kaul, P. (2020, June). Dynamic review modelling and recommendation of tourism data. In *Proceedings of the International Conference on Recent Advances in Computational Techniques (IC-RACT)*.
- Sharma, K., & Borkar, G. M. Comparative Analysis of Dynamic Web Scraping Strategies: Evaluating Techniques for Enhanced Data Acquisition, In: Ashish Kumar Tripathi and Vivek Shrivastava (eds), *Advancements in Communication and Systems*, SCRS, India, 2024, pp. 241-252. DOI: [10.56155/978-81-955020-7-3-22](https://doi.org/10.56155/978-81-955020-7-3-22)
- Stringam, B., Gerdes, J. H., & Anderson, C. K. (2023). Legal and ethical issues of collecting and using online hospitality data. *Cornell Hospitality Quarterly*, 64(1), 54-62. DOI: [10.1177/19389655211040434](https://doi.org/10.1177/19389655211040434)
- Wong, E., Rasoolimanesh, S. M., & Pahlevan Sharif, S. (2020). Using online travel agent platforms to determine factors influencing hotel guest satisfaction. *Journal of Hospitality and Tourism Technology*, 11(3), 425-445. DOI: [10.1108/JHTT-07-2019-0099](https://doi.org/10.1108/JHTT-07-2019-0099)
- Yildirim, Y., Ulucan, A., & Atici, K. B. (2023). Classifying Scuba Diving Sites through Diver Reviews with a Web Scraping Based UTADIS Application. *Croatian Operational Research Review*, 14(2), 137-148. DOI: [10.17535/crorr.2023.0012](https://doi.org/10.17535/crorr.2023.0012)
- Yuan, S. (2023). Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools. *Academic Journal of Computing & Information Science*, 6(9), 25-31. DOI: [10.25236/AJCS.2023.060904](https://doi.org/10.25236/AJCS.2023.060904)