

BIG DATA AND DATA MINING METHODS IN THE TOURISM INDUSTRY IN THE CZECH REPUBLIC

Věra Marková¹, Mikuláš Gangur¹

¹ Ing. Věra Marková*, Západočeská univerzita, Fakulta ekonomická, Katedra ekonomie a kvantitativních metod, cadkovav@students.zcu.cz, 0000-0003-3325-2899

² Doc. RNDr. Mikuláš Gangur, Ph.D., Západočeská univerzita, Fakulta ekonomická, Katedra ekonomie a kvantitativních metod, gangur@kem.zcu.cz, 0000-0002-7731-9146

Abstract: The increasing volume of data is also becoming a new challenge for the tourism industry and research. Research on the use of big data in tourism and its analysis using advanced analytical methods such as data mining, machine learning, or artificial intelligence is gaining importance worldwide. This paper aims to analyze the current research in the field of big data in tourism and tourism data processing, especially using data mining and machine learning methods, in the Czech Republic. Another aim is to compare the level of knowledge in this area in the Czech Republic and the world. The research is based on analyzing found articles dealing with this topic. These articles were analyzed in terms of the type and source of data, the analytical methods used, and the focus of the research. The results showed a slight increase in research on big data or data mining methods in tourism in the Czech Republic in recent years, but this topic is still neglected compared to the rest of the world. Researchers mostly use user-generated data, such as online hotel reviews, for sentiment analysis, fake review detection, or for classifying positive and negative reviews. A significant gap in the current research is that only a few researchers deal directly with applications in the Czech tourism environment.

Keywords: big data, data mining, tourism, review

JEL Classification: C10, C55, Z32

INTRODUCTION

In recent years the total volume of data has been growing every day. Companies from many industries can use big data to analyze and improve their business and make faster and better decisions. Big data is characterized not only by its significant volume but also by the variety of data types and the speed of its volume growth. The literature describes these characteristics as the Three V's of big data (Laney, 2001). In addition to the three V's, another dimension could also be mentioned; according to (Gandomi & Haider, 2015), they are veracity, variability, and value.

Tourism also generates big data, mainly thanks to mobile phones and the internet. Examples of big data in tourism include text data in the form of hotel or destination reviews, video or image data from social networks and websites, location information from mobile phones, or tourism business transactions (Chen et al., 2021). Choosing suitable methods to process and extract useful information is necessary for their analysis. These methods are called data mining, used to analyze and explore large amounts of data (Prasdika & Sugiantoro, 2018). Data mining is also defined as a process aimed at generating new knowledge (new patterns, relations, and trends) from data and presenting them to the user. Data mining processes apply machine learning and statistical methods to analyze data (Schuh et al., 2019). Their use in tourism could be helpful for the market analysis of possible hotel customers, understanding travelers' needs, or creating a direct mailing campaign.

(Karathiya et al., 2012). According to (Iranmanesh et al., 2022), big data analytics and artificial intelligence are emerging digital technologies in the hotel sector.

This paper's main aim is to describe the current research on tourism big data and the possibility of analyzing them, especially using data mining and machine learning methods in the Czech Republic. Another goal is to compare the level of knowledge in this area in the Czech Republic and the world, especially in the area of using big data in research, analytical method, and research focus.

1. TOURISM BIG DATA AND DATA MINING IN THE WORLD

According to Chen et al. (2021), there are three main types of big data in tourism: e-commerce data, user-generated content (UGC), and temporal-spatial behavior data. E-commerce data means data from a business transaction such as a hotel reservation or attraction ticket reservation through a website or other online platform. This data is usually not publicly available because it contains private customer information, but for entrepreneurs in the tourism industry, they represent some of the most valuable data. The user-generated content could be online tourist reviews on TripAdvisor and similar platforms or shared posts on social media such as Facebook, Twitter, Instagram, and YouTube. The temporal-spatial behavior data means data about the behavior patterns of tourists in space and time, typically GPS data.

These data sources could be users, operations, or devices (Li et al., 2018; Lyu et al., 2022; Iorio et al., 2019). According to Lyu et al. (2022), who made a content analysis of 146 articles about using big data in tourism and hospitality, most of the articles (72 %) work with user-generated data, 17 % of articles use operations data, and 10 % of them use data from devices. Only 3 % combine more than one of the sources of big data. The most common source of research using tourism big data is TripAdvisor (22 % of articles).

Advanced analytical methods, such as data mining or machine learning approaches, have been recently deployed to analyze big data in the tourism industry (Lyu et al., 2022). These methods can be used for data-driven prediction, description, or optimization (Cherian & Bhadkamkar, 2017). Predictive data mining includes classification algorithms such as decision tree methods, neural networks or support vector machines, regression, and classifier ensembles. In the tourism sector, these methods could be used for booking cancellation forecast and prediction (Antonio et al., 2019), predicting customer satisfaction (Shrestha et al., 2022), or identification of deceptive reviews (Martinez-Torres & Toral, 2019). Descriptive data mining use association rules mining as an apriori algorithm, clustering algorithms as K-means, anomaly detection techniques, and rough sets analysis. An example of optimization data mining methods could be genetic algorithms.

According to Lyu et al. (2022), most big data research in tourism and hospitality is consumer-centric. They primarily focus on tourist psychology, such as tourist behaviors, preferences about standards, or satisfaction attributes. Eight significant research focuses were identified: tourist psychology, big data and artificial intelligence, mobility pattern, conventional tourism, tourism marketing, destination management, Electronic Word of Mouth (eWOM) management, and big data and knowledge management. Chen et al. (2021) defined the main domains where are tourism big data applied as assisting smart tourism development (smart procedure management, customer image capturing, market trend analysis), measuring psychological contents of customers (sentiment or satisfaction), and content innovation via big data (forecast market demand and improve their service).

2. TOURISM BIG DATA AND DATA MINING IN THE CZECH REPUBLIC

2.1 Methodology

For a systematic review of the published works related to the Czech Republic in the research of tourism, big data and data mining methods (possibly machine learning or artificial intelligence) were used databases Scopus, Web of Science (WoS), and Google Scholar. As publications related to the Czech Republic,

publications written by Czech authors (the main or second author in the order has an affiliation in the Czech Republic) or publications that describe the application of data mining methods on Czech data were selected. Table 1 shows the search strategy used in this research. In the case of the Scopus and WoS databases, in the next step, the results were limited by country to the Czech Republic or by adding the word "czech" to the string of search criterium. In the case of Google Scholar, this was limited directly at the first search stage and subsequently selected according to the authors' affiliations or data sources. We searched for publications without publication date restrictions, i.e., all publications published on this topic by August 2022.

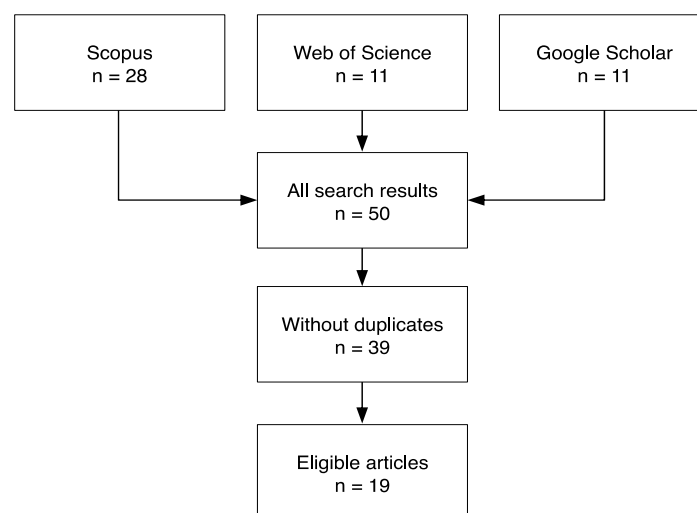
Tab. 1: Search strategy

Databases	Search Criteria	Search fields
Scopus	(„data mining“ OR "big data" OR "machine learning" OR "Artificial Intelligence") AND (tourism OR hospitality OR hotel))	abstract, title, keywords
Web of Science	(„data mining“ OR "big data“ OR "machine learning“ OR "Artificial Intelligence") AND (tourism OR hospitality OR hotel))	abstract, title, keywords
Google Scholar	("data mining" OR "big data" OR "machine learning" OR "Artificial Intelligence")(tourism OR hospitality OR hotel) ("czech republic" OR czech)	all fields

Source: own processing

The selection process of the publications continued by reading the titles and abstracts and, if necessary, the full article. Based on this, duplicates were excluded, and publications suitable for more detailed analysis were selected. Eligibility of publications was assessed regarding the main topic (big data/data mining in tourism) and according to the connection with the Czech Republic. Figure 1 shows the whole process of publication searching. At the end of the process, we obtained 19 relevant publications for further analysis; 16 articles were found in full text, three only as abstracts. There are two reviews and 17 application articles; 18 are by Czech authors, and 1 used only Czech data.

Fig. 1: Article selection process

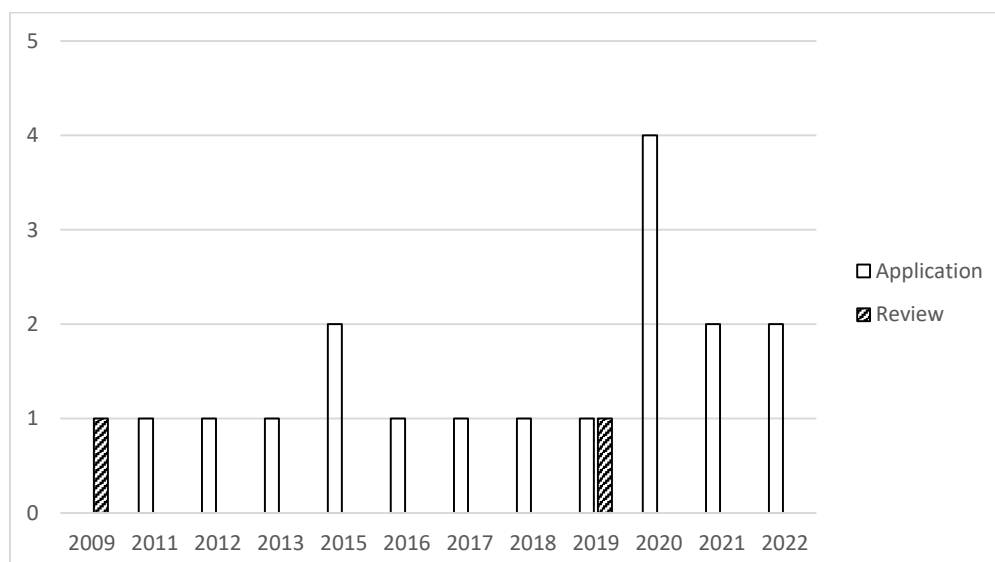


Source: own processing

2.2 Results

First Czech article that deals with data mining topics in tourism was published in 2009 by Čech and Bureš (2009). The authors described in their review article advanced technologies in e-tourism, one part of the article was focused on data mining. The application of data mining in tourism has been rare as of this article's publication date. Still, the authors assumed that data mining and other advanced technologies would extend in the future due to the increasing volume of data and information. The second review article by Czech authors was published in 2019 (Emmer & Holešinská, 2019). This review article describes the sources and types of big data in tourism and their potential application in destination management. The authors offer an overview of big data application research cases depending on the data sources used and provide an overview of the tools and methods used for the big data processing. Since 2011, Czech authors mainly published one article on a topic related to applications of big data or data mining in tourism per year until 2020, when their number began to grow, see Fig. 2. In 2021 two articles were published on this topic as well as in January - August 2022.

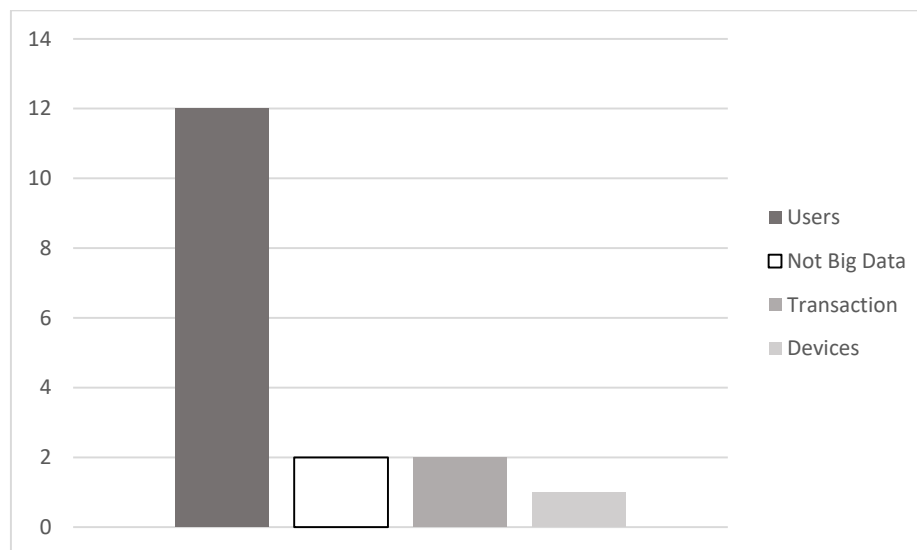
Fig. 2: Articles by year (data in 2022 is until the end of August)



Source: own processing

Figure 3 shows a count of published articles divided by the source of used data. Only those articles that described the application of data mining methods or the use of big data are displayed, i.e., a total of 17 articles. The most common source of big data in the publications was users; two articles used transaction data, and only one used data from devices. Two papers were used for data mining applications, not big data. User data used in found articles mainly were (11 articles) hotel or restaurant reviews from Booking, TripAdvisor, or similar web sources. All these articles deal with implementing new approaches and methods to analyze these reviews, primarily focused on sentiment analysis, fake review detection, or identifying positive/ negative reviews. One article analyzed social media posts on Facebook (Hruska, 2020).

Fig. 3: Source of big data used in Czech tourism research (only application articles, n = 17)



Source: own processing

The first publication that used user-generated data for research in the Czech Republic was by Dařena and Žižka (2011). They described the method of classifying selected words from hotel reviews into those with positive or negative meaning using trained decision trees. Žižka et al. (2012) deal with almost two million hotel service reviews in English in the following article. Their experiment aimed to evaluate the options of clustering algorithms for automatically categorizing reviews into positive or negative. Dařena & Žižka (2013) presented a new alternative method to standard clustering techniques named Semi-supervised Classification (SuDoC). Žižka and Svoboda (2015) describe the possibility of using windowing, a technique for generating subsamples of a data set that provide enough information to train a model for classification and obtain results similar to those achieved by training a model on the entire data set. For the classification of words from hotel reviews (positive/ negative), the decision tree c5 was used. Based on the experiments, they suggested a suitable training-set size estimation method. In the following article, Žižka & Dařena (2015) deal with sentiment analysis and options of possible automatized methods of generating relevant phrases using decision trees. The online hotel reviews in 4 languages (English, German, Spanish, and Russian) were used. The authors tested this method on a dataset containing 1245 positive and negative hotel reviews. They compared it with the results of other techniques such as Support Vector Machines, Decision Trees, Logistic Regression, and others. Hán et al. (2016) dealt with using modern analytical software with data mining and text mining tools to analyze hotel reviews. Other authors, Vencovský et al. (2017), compare four text mining techniques for extracting service aspects. To do this, they used a dataset that contains 1 274 reviews collected from an unspecified online review site. Barusha and Hájek (2019) discussed in their paper the possibility of review spam detection and presented a new content-based approach using a deep feed-forward neural network for classification. The authors used datasets from Cornell University with positive and negative hotel reviews containing legitimate and spam reviews. Results of their method were compared with other methods such as Convolutional neural network, Support Vector Machine, Random Forest, and Naive Bayes classifier. Hájek et al. (2021) presented the new sentiment analysis model that uses emotion associations and explained topic models in another article. The authors used Amazon product reviews and hotel reviews for their experiment. Hájek and Sahut (2022) used a large benchmark dataset of Yelp restaurant reviews to detect fake reviews. The dataset contains 608 598 reviews for 5 044 restaurants in selected states of the USA posted by 260 277

reviewers. In their experiment, they constructed the machine learning model based on a neural network to evaluate the effectiveness of the sentiment-dependent linguistic features for fake review detection.

The only article included in this analysis that does not come from Czech authors is by Nilashi et al. (2021), which describes the method of predicting travelers' preference for eco-friendly hotels using a combination of supervised and unsupervised machine learning algorithms. The experiments were performed on TripAdvisor's collected data from the Czech Republic's eco-friendly hotel.

The only author who used data other than hotel reviews was Hruška (2020). He used Facebook posts of the National Tourism Organization (NTOs) from 10 countries (France, Turkey, China, Germany, Australia, Italy, Mexico, USA, Spain, and Great Britain) to make a sentiment analysis. His purpose was to compare the sentiment of NTOs posts with follower growth. For sentiment analysis was used a Google product and library package Artificial Intelligence and Machine Learning with Google Natural Language API.

Two articles describe the use of transaction data and come from the same authors; both deal with market segmentation (Chalupa & Petricek, 2020; Chalupa & Petricek, 2022). In the first paper from 2020, the authors described two-step clustering using data from online booking platforms such as Booking.com and others. This study's main focus was on understanding hotel guests booking behavior to planning marketing activities and revenue marketing optimization. Five variables were selected for clustering: the room type name, rate plan, net average room om rate, rate, room rate, booking window, and the number of room nights. Based on this, 4 clusters of customers were identified. In the following article from 2022, the authors used the transaction data of four-star hotels located in Prague (9 485 transactions) for market segmentation. The final 6 clusters were provided by the K-means clustering and Two-Step cluster analysis. The authors also measured the coefficient of price elasticity for these segments for possible price optimization.

The only article that used data from devices was by Pavlicek et al. (2018). This article describes the possibility of analyzing geolocation big data originating from a mobile network operator in the Czech Republic. The authors present four case studies: National Park visitors' distribution analysis, analysis of the behavior of visitors to Czech mountain resorts (combination with another source of data – surveys), data analysis for the preparation of city territorial and development plan and using this data to efficient tourism management at Vaclav Havel Airport. Unfortunately, the article did not describe the data analysis methods in detail.

Two articles included in this review did not work with big data but used data mining methods for small data analysis. Botlíková et al. (2020) use the K-means algorithm for cluster districts of the Czech Republic to identify those with high tourism intensity and to demonstrate the relationship between tourism intensity (using Defert's function and modified functions) and socio-economic factors (as criminality or migration) using relational analysis. The other article by Hašková and Horák (2020) uses the fuzzy logic approach to evaluate the degree of country travel and tourism competitiveness. As data, they used the subjective rankings of Travel & Tourism Policy and Enabling Conditions, Natural and Cultural Resources, and Safety and Security from the perception of a traveling person. They evaluated five countries: the Czech Republic, China, Pakistan, Russia, and Turkey.

In Table 2, there is a summary of knowledge obtained from the analysis of reviewed articles. The first column represents the data source, and the other three columns contain a specific example of data using analytical methods and research focus from analyzed articles.

Tab. 2: Summary of article analysis

Data source	Example of data	Used analytical methods	Research focus	Author
Users	Hotel reviews (Booking.com etc.), social media posts	K-means, Decision Trees, Random Forest, Neural Networks, Support Vector Machines, Logistic Regression, Naive Bayes Classifier	Identifying of positive or negative reviews, spam review detection, sentiment analysis	Žižka (2011), Žižka et al. (2012), Dařena & Žižka (2013), Žižka & Svoboda (2015), Žižka & Dařena (2015), Han et al. (2016), Vencovsky et al. (2017), Barusha & Hajek (2019), Hruřka (2020), Hajek et al. (2021), Hajek & Sahut (2022), Nilashi et al. (2021)
Transaction	Data from the booking transaction (room type, room rate etc.)	Two-step clustering, K-means	Market segmentation	Chalupa & Petricek (2020), Chalupa & Petricek (2022)
Devices	Geolocation data	Business Intelligence	Visitors distribution analysis, analysis of the visitors behavior, efficient tourism management	Pavlicek et al. (2018)

Source: own processing

CONCLUSION

The main aim of this article was to describe the current research on tourism big data and the possibility of analyzing them, especially by using data mining and machine learning methods in the Czech Republic. For this purpose, 19 articles from this area written by Czech authors (18 articles) or whose authors used data from the Czech environment for research (1 article) were analyzed. Emphasis was placed on the type of big data used, methods for analyzing this data, and research focus.

According to authors worldwide, big data and data mining methods in tourism and hospitality have promising potential, and research on this topic is growing (Chen et al., 2021; Rahmadian et al., 2021; Lyu et al., 2022). The study of world trends in this area, described in the second chapter of that article, shows that researchers work mainly with user-generated big data, less research describes the analysis of operational data, and the least use is data from devices. Related to this are trends in research focus, most of the big data research in tourism and hospitality in the consumer domain. Due to the diverse tasks that can be solved in tourism and hospitality, a wide range of data mining methods can be used to analyze big data.

A deeper analysis of 19 articles from the Czech environment in the third chapter of this article showed the current state of research on this topic in the Czech Republic. Although the number of publications has slightly increased in recent years, researchers in the Czech Republic are not as interested in this topic as elsewhere. Research by Czech authors mainly deals with text mining tasks and analysis of hotel reviews as sentiment analysis or spam review detection, and only a few of them solve applications directly in the Czech environment (with data related to the Czech Republic). As in the world, Czech authors use most of the users' big data in

their research. Due to the small number of found articles, it is impossible to determine the most frequently used data mining method precisely.

This article is the first to create an overview of current research on big data and data mining methods in the Czech Republic. It creates opportunities to find new research topics for authors from the Czech Republic. The biggest gap is the lack of applications built on big data from the Czech environment, i.e., from Czech hotels, restaurants, or destination agencies. The question here is whether these subjects are interested in this type of research. All businesses in the tourism and hospitality industry automatically collect data about their clients, whether through bookings, social media posts, or hotel reviews, one of the future research goals could be to map how these subjects work with this data and find out the needs of these subjects in the field of big data analysis. Considering the possibilities of using big data for managerial decision-making in the tourism industry, it would also be interesting to focus on using business intelligence in this area.

Acknowledgment

This paper was created with the subsidy of the project SGS-2021-022 Financial (stock) markets, modeling and prediction of behavior carried out with the support of the University of West Bohemia.

REFERENCES

- Antonio, N., de Almeida, A., & Nunes, L. (2019). Big Data in Hotel Revenue Management: Exploring Cancellation Drivers to Gain Insights Into Booking Cancellation Behavior. *Cornell Hospitality Quarterly*, 60(4), 298-319. <https://doi.org/10.1177/1938965519851466>
- Barushka, A., & Hajek, P. (2019). Review Spam Detection Using Word Embeddings and Deep Neural Networks. *Artificial Intelligence Applications and Innovations*, 340-350. https://doi.org/10.1007/978-3-030-19823-7_28
- Botlíková, M., Botlík, J., & Janík, P. (2020). Hodnocení cestovního ruchu v okresech České republiky pomocí strojového učení. In *Aktuální problémy cestovního ruchu - Overturismus – Riziko pro destinace. 15th Conference* (pp. 11-20). Vysoká škola podnikání Jihlava.
- Chalupa, S., & Petricek, M. (2020). Using Customer Characteristics to Manage Marketing and Revenue Management Activities. *TEM Journal*, 1088-1093. <https://doi.org/10.18421/TEM93-33>
- Chalupa, S., & Petricek, M. (2022). Understanding customer's online booking intentions using hotel big data analysis. *Journal of Vacation Marketing*. <https://doi.org/10.1177/13567667221122107>
- Cherian, T., & Bhadkamkar, H. A study and survey of big data using data mining techniques. *Journal of Engineering Sciences & Research Technology*, 6(10), 169-174. <https://doi.org/10.5281/zenodo.1002693>
- Chen, H., Liu, Y., & Chen, K. (2021). Big Data in Tourism: General Issues and Challenges. *J Tourism Hospit*, 10(4). <https://doi.org/10.35248/2167-0269.21.s4.003>
- Čech, P., & Bureš, V. (2009). Advanced Technologies in e-Tourism. In *Proceedings of the 9th WSEAS International Conference on APPLIED COMPUTER SCIENCE* (pp. 85-92).
- Dařena, F., & Žižka, J. (2011). Text mining-based formation of dictionaries expressing opinions in natural languages. *Proceedings of the 17th International Conference on Soft Computing Mendel 2011*, 374-381.
- Dařena, F., & Žižka, J. (2013). SuDoC: Semi-supervised Classification of Text Document Opinions Using a Few Labeled Examples and Clustering. *FQAS 2013*, 625-636.
- Emmer, F., & Holešinská, A. (2019). Big Data: a Source of Mobility Behaviour and a Strategic Tool for Destination Management. *Czech Journal of Tourism*, 8(2), 85-102. <https://doi.org/10.2478/cjot-2019-0006>

- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, (35), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Hájek, P., Barushka, A., & Munk, M. (2021). Neural Networks with Emotion Associations, Topic Modeling and Supervised Term Weighting for Sentiment Analysis. *International Journal of Neural Systems*, 31(10). <https://doi.org/10.1142/S0129065721500131>
- Hájek, P., & Sahut, J. -M. (2022). Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection. *Technological Forecasting and Social Change*, 177. <https://doi.org/10.1016/j.techfore.2022.121532>
- Hán, J., Lustigová, Z., & Chalupa, Š. (2016). Data mining in hospitality. In . Slezská univerzita v Opavě.
- Hašková, S., & Horák, J. (2020). Fuzzy logic and artificial intelligence tools in the context of tourism analysis. In *5th International Thematic Monograph: Modern Management Tools and Economy of Tourism Sector in Present Era* (pp. 601-617).
- Hruška, J. (2020). Sentiment Analysis of National Tourism Organizations on Social Media. *Hradec Economic Days*, 250-256. <https://doi.org/10.36689/uhk/hed/2020-01-028>
- lorio, C., Pandolfo, G., D'Ambrosio, A., & Siciliano, R. (2020). Mining big data in tourism, 54(5-6), 1655-1669. <https://doi.org/10.1007/s11135-019-00927-0>
- Iranmanesh, M., Ghobakhloo, M., Nilashi, M., Tseng, M. -L., Yadegaridehkordi, E., & Leung, N. (2022). Applications of disruptive digital technologies in hotel industry: A systematic review. *International Journal of Hospitality Management*, 107. <https://doi.org/10.1016/j.ijhm.2022.103304>
- Karathiya, M., Sakshi, R. S., Sakshi, D. S., & Karathiya, D. (2012). Data mining for travels and tourism. *Journal of Information and Operations Management*, 3(1), 114-118.
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323. <https://doi.org/10.1016/j.tourman.2018.03.009>
- Lyu, J., Khan, A., Bibi, S., Chan, J. H., & Qi, X. (2022). Big data in action: An overview of big data studies in tourism and hospitality literature. *Journal of Hospitality and Tourism Management*, 51, 346-360. <https://doi.org/10.1016/j.jhtm.2022.03.014>
- Martinez-Torres, M. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75, 393-403. <https://doi.org/10.1016/j.tourman.2019.06.003>
- Nilashi, M., Minaei-Bidgoli, B., Alrizq, M., Alghamdi, A., Alsulami, A. A., Samad, S., & Mohd, S. (2021). An analytical approach for big social data analysis for customer decision-making in eco-friendly hotels. *Expert Systems with Applications*, 186. <https://doi.org/10.1016/j.eswa.2021.115722>
- Pavlicek, A., Doucek, P., Novák, R., & Strizova, V. (2018). Big Data Analytics – Geolocation from the Perspective of Mobile Network Operator. *Research and Practical Issues of Enterprise Information Systems*, 119-131. https://doi.org/10.1007/978-3-319-94845-4_11
- Prasdika, P., & Sugiantoro, B. (2018). A Review Paper on Big Data and Data Mining: Concepts and Techniques. *International Journal on Informatics for Development*, 7(1), 36-38. <https://doi.org/10.14421/ijid.2018.07107>
- Rahmadian, E., Feitosa, D., & Zwitter, A. (2022). A systematic literature review on the use of big data for sustainable tourism. *Current Issues in Tourism*, 25(11), 1711-1730. <https://doi.org/10.1080/13683500.2021.1974358>

- Schuh, G., Reinhart, G., Prote, J. -P., Sauermann, F., Horsthofer, J., Oppolzer, F., & Knoll, D. (2019). Data Mining Definitions and Applications for the Management of Production Complexity. *Procedia CIRP*, 81, 874-879. <https://doi.org/10.1016/j.procir.2019.03.217>
- Shrestha, D., Wenan, T., Gaudel, B., Rajkarnikar, N., & Jeong, S. R. (2022). Multi-dimensional Analysis and Prediction Model for Tourist Satisfaction. *KSII Transactions on Internet and Information Systems*, 16(2), 480-502. <https://doi.org/10.3837/tiis.2022.02.007>
- Vencovský, F., Lucas, B., Mahr, D., & Lemmik, J. (2017). Comparison of Text Mining Techniques for Service Aspect Extraction. *4th European Conference on Social Media*.
- Žižka, J., & Dařena, F. (2015). Automated Mining of Relevant N-grams in Relation to Predominant Topics of Text Documents. *TSD 2015*, 461-460. https://doi.org/10.1007/978-3-319-24033-6_52
- Žižka, J., & Svoboda, A. (2015). Customers' Opinion Mining from Extensive Amount of Textual Reviews in Relation to Induced Knowledge Growth. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 63(6), 2229-2237. <https://doi.org/10.11118/actaun201563062229>
- Žižka, J., Burda, K., & Dařena, F. (2012). Clustering a Very Large Number of Textual Unstructured Customers' Reviews in English. *AIMSA 2012*, 38-47.