PAPER NAME

Big Data in Tourism Destinations A Systematic Literature Review.pdf

AUTHOR

ERNI WIDARTI

WORD COUNT

3749 Words

CHARACTER COUNT

20517 Characters

PAGE COUNT

6 Pages

FILE SIZE

183.0KB

SUBMISSION DATE

May 7, 2024 2:55 PM GMT+7

REPORT DATE

May 7, 2024 2:55 PM GMT+7

7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 4% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 0% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 10 words)
- Manually excluded sources

Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 4 Iss. 3 July 2024, pp: 719-724
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Big Data in Tourism Destinations: A Systematic Literature Review

Erni Widarti[1*], Moh. Erkamim[2], Wartono[3]

[1,2,3]Program Studi Sistem Informasi Kota Cerdas, Fakultas Teknik, Universitas Tunas Pembangunan Surakarta (UTP), Indonesia

E-Mail: [1]erni.widarti@lecture.utp.ac.id, [2]erkamim@lecture.utp.ac.id, [3]wartono@lecture.utp.ac.id

Received Jan 05th 2024; Revised May 08th 2024; Accepted Jun 10th 2024
Corresponding Author: Erni Widarti

Abstract

Tourism is an industrial sector that has a variety of data originating from tourists. This data can be utilized and reprocessed through the application of artificial intelligence technology such as big data and machine learning to analyze and predict tourist patterns so that it can be used in developing the tourism sector. This research aims to present a systematic literature review regarding the role of big data and machine learning in the tourism context. This research reviews 25 research papers related to big data and machine learning applied in the tourism industry. The categorization of tourism research related to big data and machine learning is based on research published from 2018 to 2023. The focus of this research is to provide an in-depth review based on journal rankings, research objectives, types of data used, and algorithms applied in the research. The results of this research are to provide a systematic literature review that can be used to help future researchers discover new research topics and present insights into future prospects regarding the use of big data and machine learning in tourism.

Keyword: Big Data, Destinations, Systematic Literature Review, Tourism

1. INTRODUCTION

Tourism is a recreational trip or vacation to a place that has an attraction for visitors. The attraction can be in the form of having historical values, ancestral heritage, natural beauty, etc., with tourist attraction can increase tourists who come and visit. The tourism industry is one sector that has an increase every year, for example in Indonesia. According to the Central Statistics Agency in Indonesia, it was stated that from January to June 2018, it was stated that 7.53 million foreign tourists visited Indonesia. The number of visits rose to 13.08 percent from the number of visits in 2017 which amounted to 6.66 million tourists [1]. The increase in the number of tourists produces large data. One of the data generated from the tourism industry comes from tourists. The data can be processed for use in the development of the tourism industry. Data relating to tourism is divided into three main categories, namely 1) user data generated by tourists, 2) device data generated by cellular devices, 3) transaction data generated by online transactions such as online order [2].

The tourism industry is undergoing a major transformation driven by developments in digital technology and the availability of large amounts of data. This transformation creates opportunities and challenges for tourism destinations that want to increase competitiveness and sustainability. Big data analysis has an important role in the tourism industry, especially in improving and optimizing the operational efficiency of tourism destinations [3]. By analyzing data produced by tourism destinations, it can be processed properly to understand trends in tourist behavior, tourist preferences and travel patterns which can improve the sustainability of tourism destination development [4].

Problems faced in using big data analysis in tourism destinations include challenges in collecting, managing and analyzing complex data. Therefore, this research focuses on big data methods used in processing data generated in the tourism sector. Big data has an important role in determining tourism businesses to formulate strategies, business tactics, and policies in tourism companies. Data generated from tourists can be used for Smart Tourism purposes [5]. The results of this dataset can be used to monitor the flow of tourists, analyze [6] and predict traveler travel behavior based on digital traces [7]. Big Data supports business leaders and HR managers in developing clear strategies for the right acquisition and development of skills [8].

The aim of this reseach is to provide a review of the implementation of big data and machine learning in the area of tourism. The focus of this study is to review based on journal rankings, the objectives of the research study, the data used, and the algorithms used. The urgency of this research lies in the need to understand the role and potential of big data in the tourism context, as well as identifying future research

directions that can increase the use of big data for the sustainability of tourism destinations in the era of digital transformation. Thus, it is hoped that this research can provide valuable insights for practitioners, researchers and policy makers in the tourism sector.

2. RELATED WORKS

The rapid development of the tourism industry and the increasing number of tourists, produce large data. Some research utilizes the big data method to process data produced by the tourism industry [9]. Based on Jingjing Li's research study, a major contribution has been summarized into three aspects, namely reviewing the scale of the kinds of large data used in tourism study, analyzing various types of large data from the perspective of research focus (on tourism problems), data features and analytical methods, conducting comprehensive surveys, provide information about new issues and future prospects for big data in tourism [2].

Based on M. Mariani's research, conducted a literature review with the aim of identifying future research and development gaps in designing an agenda for future research. This study consists of literature reviews based on the following parameters: topics of research; characterization base on conceptual and theoretical; data source of research; size and type of data; method of collecting data; data analysis technique; reporting and visualization of data. The result is a literature review can helps identify business problems, integration between business intelligence and big data domains with management in the development of tourism and hotels that are not yet available [10]. Some of the uses of big data in tourism include quantifying nature-based tourism in protected areas [11], Classifying multi-destination trips[12], estimate the arrival of tourist destinations [13], analyze tourism patterns [14], sentiment analysis in tourism [15], used in smart tourist destinations [5], analyze tourist behavior [6]. From a number of reviews, the research can be used to identify research gaps, develop future research and design or determine topics for future research [10].

The originality of this study is to current a literature review of the usefulness of big data in tourism by identifying and analyzing existing research studies between 2018-2023 based on the purpose of the study, the data used and the algorithms used in the study. The result from literature review can be used to identify and analyze research gaps, develop future research, predict tourism behavior and design or determine innovative topics for future research and provide marketing strategies tourism to smart tourism purpose.

3. METHOD

This study is a quantitative literature review of previous research studies on large data in tourism. The method used is Systematic Literature Review (SLR), which is a systematic observation of literature that aims to identify, analyze, generate and interpret the findings contained in the research. The research studies chosen were between 2018 and 2023. The study was taken from several websites that published research papers, which are international journals. After that, specify the journal quartile ranking based on the journal quartile index rank on Scimago web [16]. Then categorize previous research based on the objectives of the research study, the data used, and the algorithms used in the study. A flow diagram of research methodologies can be seen in Figure 1.

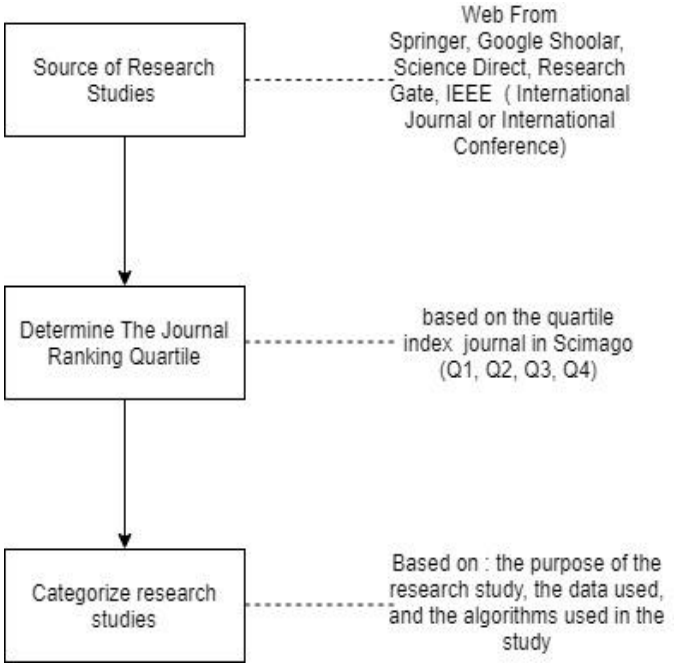

Figure 1. Research Methodology

Figure 1 describes the steps taken in the methodology. the first stage is to look for research studies that have been published from 2018 until 2023. The website used is Springer, Google Scholar, Science Direct, Research Gate, IEEE. The rationale for using the selected sources, such as Springer, Google Scholar, Science Direct, Research Gate, and IEEE, lies in their reputation and credibility within the academic. These sources are widely recognized for their extensive coverage of scholarly literature in various disciplines, including tourism and big data. Springer, Science Direct, and IEEE are known for their journals and conference proceedings, which are valuable sources of academic research. Google Scholar is provides a comprehensive search of scholarly articles, theses, books, and conference papers, making it a valuable tool for conducting a thorough literature review. Research Gate is a platform where researchers can share their work and collaborate with others, making it a useful resource for accessing the latest research in the field. The next step is to determine the ranking quartile index based on the journal quartile index on the Scimago website [16], namely Q1, Q2, Q3, Q4. Furthermore, it categorizes based on the purpose of the research study, the data used and the algorithms used in the study.

4. RESULTS AND DISCUSSION

At this stage, the 25-research study chosen must be in the big data topic area in tourism. Previous studies were taken from several websites that published research papers, which are international journals or international conferences. The focus of this study is to review based on journal rankings, the objectives of the research study, the data used, and the algorithms used in the study. How to determine the journal quartile index rank based on the journal quartile index on the Scimago website [16], namely Q1, Q2, Q3, Q4. Table 1 is the result of a paper that has been known to rank its journal quartile.

Table 1. Results from Journal Ranking

Title	Journal Quartile Ranking	Citation
Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model	Q1	[13]
Creating value from Social Big Data: Implications for Smart Tourism Destinations	Q1	[5]
Sentiment Analysis in Tourism: Capitalizing on Big Data	Q1	[15]
Tourists' digital footprint in cities: Comparing Big Data sources.	Q1	[7]
Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources.	Q1	[14]
Forecasting tourist arrivals with machine learning and internet search index	Q1	[17]
Classifying multi-destination trips in Austria with big data.	Q1	[12]
Quantifying nature-based tourism in protected areas in developing countries by using social big data.	Q1	[11]
'You will like it!' using open data to predict tourists' response to a tourist attraction	Q1	[18]
Forecasting tourism demand with composite search index	Q1	[19]
Forecasting accuracy evaluation of tourist arrivals	Q1	[20]
A survey towards an integration of big data analytics to big insights for value-creation	Q1	[4]
The concept of smart tourism in the context of tourism information services	Q1	[3]
What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management.	Q1	[21]
Hybrid Recommender System for Tourism Based on Big Data and AI: A Conceptual Framework	Q1	[22]
New avenues for second home tourism research using big data: prospects and challenges	Q1	[23]
Tourism scene classification based on multi-stage transfer learning model	Q2	[24]
Applying Big Data Analytics to Monitor Tourist Flow for the Scenic Area Operation Management.	Q3	[6]
Tourism Recommendation Using Machine Learning Approach	Q3	[25]

Based on table 1, it can be seen that the journal quartile index ranking is based on scimago. The results of determining this index ranking can be used to determine the quality of the paper based on the assessment of the Scimago website [16]. After determining the journal ranking, then categorizing the paper based on the objectives of the study regarding big data in the arena of tourism. The method of big data and machine learning

can be used in various purposes of research on tourism in various ways such as predicting, analyzing, identifying, as in Table 2 which explains the purpose of big data use in tourism.

Table 2. The Purpose of Big Data in Tourism

Aim/purpose	Citation
Big data is used for innovation Smart Tourism	[5][3][23][26] [27][28][29]
Detect / predict for trips tourism	[30][18][17][19][20]
Recommend tourism trips	[25][15][22][31][32]
Analyzing the digital footprint of tourists using Big Data	[7] [6][21]
Classify for tourism trips	[12][24]
Analyze tourism trips patterns	[14]
Explore the correlation of weather, temperature, weekends and national holidays to estimate the arrival of tourist destinations.	[13]
Quantifying natural tourism with big data	[11]

Based on table 2 states that big data plays a function in the tourism sector with a variety of different objectives. Big data can be used in developing and innovation smart tourism, analyzing digital footprints, analyzing patterns, analyzing tourist travel behavior, estimating tourist arrivals, classifying multi-purpose trips, etc. The outcomes of the use of the big data method are used for data development and processing in the tourism sector. In addition, it can also be used to offer new research in tourism in the future.

Furthermore, grouping based on the data used in research on big data in research. Data grouping is based on previous research which categorizes data into 3 are UGC data, device data, and transaction data [2]. In addition, one category is added, namely descriptive data such as weather, climate, natural conditions. Broadly speaking, grouping data as in table 3.

Table 3. Data categories

Data Category	Citation
Transaction Data	[14][11][32][18][17][19][20] [27][21][22]
UGC data (data generated by the user)	[12][5][15][31][30][25][26] [28]
Device Data	[7][6][24][3][23]
Descriptive data	[13]

Based on table 3, the data that is used the most is transaction data. Transaction data is based on operations including web data search, web page visit data, online ordering data. Some journals use transaction data because data taken online from history or from the web is easier to obtain. The second rating is data from users, such as photos or social media used by users.

After grouping based on the data used, then grouping the paper based on the model/technique/ algorithm used in the study. There are several types of models used in table 4.

Table 4. Algorithm / Model / Technique used

Algorithm / Model / Technique	Citation
Naïve Bayes Classifiers	[12][30][3] [32] [7]
Kernel extreme learning machine	[17][26] [18]
Deep-learning based sentiment analysis	[21] [15]
Convolutional Neural Network (CNN)	[24] [27]
Keyhole dan Buzztrack	[5][28]
Collaborative filtering	[31]
Autoregressive Vector Modeling	[13]
Regression	[25]
Generalized dynamic factor model	[19]
Resilient Distributed Dataset (RDD)	[6]
Natcap.invest.recreation from Natural Capital Project.	[11]
Recurrent Singular Spectrum Analysis model	[20]

Based on table 4 algorithms/models/ Techniques that are widely used are Naïve Bayes Classifiers, extreme learning machine kernels, and analytical sentiment. The algorithm is widely used in big data and machine learning in tourism to predict, recommend, smart tourism, analysis, and classification of tourist patterns.

4. CONCLUSION

From the results of the review, it was concluded that the presence of big data and machine learning can usage in the processing and developing datasets produced by tourists. The big data and machine learning method can usage to monitor the flow of tourists, analyze and predict behavior, recommendations, smart tourism, classify travel, etc.

The result of this research review is to provide review literature that can be used to help future researchers find new research topics and present valuable insights about future prospects about the usage of big data and machine learning in tourism. In further research, it can be further developed by categorizing based on the applications used, for example, Weka, Google Colab, etc. Besides that, it can be developed not only in the tourism sector but also in the hotel sector because it is interconnected.

REFERENCES

[1] P. Pariwisata, "STATISTIK Perkembangan Pariwisata dan Transportasi Nasional 2019," no. 13, pp. 1–12, 2019.

[2] J. Li, L. Xu, L. Tang, S. Wang, and L. Li, "Big data in tourism research : A literature review," *Tour. Manag.*, vol. 68, pp. 301–323, 2018, doi: 10.1016/j.tourman.2018.03.009.

[3] Y. Li, C. Hu, C. Huang, and L. Duan, "The concept of smart tourism in the context of tourism information services," *Tour. Manag.*, vol. 58, pp. 293–300, 2017, doi: 10.1016/j.tourman.2016.03.014.

[4] M. K. Saggi and S. Jain, "A survey towards an integration of big data analytics to big insights for value-creation," *Inf. Process. Manag.*, vol. 54, no. 5, pp. 758–790, 2018, doi: 10.1016/j.ipm.2018.01.010.

[5] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from Social Big Data: Implications for Smart Tourism Destinations," *Inf. Process. Manag.*, vol. 54, no. 5, pp. 847–860, 2018, doi: 10.1016/j.ipm.2017.10.006.

[6] S. Qin, J. Man, X. Wang, C. Li, H. Dong, and X. Ge, "Applying Big Data Analytics to Monitor Tourist Flow for the Scenic Area Operation Management," *Discret. Dyn. Nat. Soc.*, vol. 2019, pp. 1–11, 2019, doi: 10.1155/2019/8239047.

[7] M. H. Salas-Olmedo, B. Moya-Gómez, J. C. García-Palomares, and J. Gutiérrez, "Tourists' digital footprint in cities: Comparing Big Data sources," *Tour. Manag.*, vol. 66, pp. 13–25, 2018, doi: 10.1016/j.tourman.2017.11.001.

[8] S. Verma, S. S. Bhattacharyya, and S. Kumar, "An extension of the technology acceptance model in the big data analytics system implementation environment," *Inf. Process. Manag.*, vol. 54, no. 5, pp. 791–806, 2018, doi: 10.1016/j.ipm.2018.01.004.

[9] C. McCamley and A. Gilmore, "Aggravated fragmentation: A case study of SME behaviour in two emerging heritage tourism regions," *Tour. Manag.*, vol. 60, pp. 81–91, 2017, doi: 10.1016/j.tourman.2016.11.016.

[10] M. Mariani, R. Baggio, M. Fuchs, and W. Höepken, "Business intelligence and big data in hospitality and tourism: a systematic literature review," *Int. J. Contemp. Hosp. Manag.*, vol. 30, no. 12, pp. 3514–3554, 2018, doi: 10.1108/IJCHM-07-2017-0461.

[11] Y. Kim, C. ki Kim, D. K. Lee, H. woo Lee, and R. I. T. Andrada, "Quantifying nature-based tourism in protected areas in developing countries by using social big data," *Tour. Manag.*, vol. 72, no. February 2018, pp. 249–256, 2019, doi: 10.1016/j.tourman.2018.12.005.

[12] I. Önder, "Classifying multi-destination trips in Austria with big data," *Tour. Manag. Perspect.*, vol. 21, pp. 54–58, 2017, doi: 10.1016/j.tmp.2016.11.002.

[13] Y. Y. Liu, F. M. Tseng, and Y. H. Tseng, "Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model," *Technol. Forecast. Soc. Change*, vol. 130, no. February 2017, pp. 123–134, 2018, doi: 10.1016/j.techfore.2018.01.018.

[14] F. Batista e Silva, M. A. Marín Herrera, K. Rosina, R. Ribeiro Barranco, S. Freire, and M. Schiavina, "Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources," *Tour. Manag.*, vol. 68, pp. 101–115, 2018, doi: 10.1016/j.tourman.2018.02.020.

[15] A. R. Alaei, S. Becken, and B. Stantic, "Sentiment Analysis in Tourism: Capitalizing on Big Data," *J. Travel Res.*, vol. 58, no. 2, pp. 175–191, 2019, doi: 10.1177/0047287517747753.

[16] Scimagojr, "Scimago Journal & Country Rank https://www.scimagojr.com/." https://www.scimagojr.com/journalsearch.php (accessed Oct. 10, 2018).

[17] S. Sun, Y. Wei, K. L. Tsui, and S. Wang, "Forecasting tourist arrivals with machine learning and internet search index," *Tour. Manag.*, vol. 70, no. February 2018, pp. 1–10, 2019, doi: 10.1016/j.tourman.2018.07.010.

[18] E. Pantano, C. V. Priporas, and N. Stylos, "'You will like it!' using open data to predict tourists' response to a tourist attraction," *Tour. Manag.*, vol. 60, pp. 430–438, 2017, doi: 10.1016/j.tourman.2016.12.020.

[19] X. Li, B. Pan, R. Law, and X. Huang, "Forecasting tourism demand with composite search index," *Tour.*

Manag., vol. 59, pp. 57–66, 2017, doi: 10.1016/j.tourman.2016.07.005.
[20] H. Hassani, E. S. Silva, N. Antonakakis, G. Filis, and R. Gupta, "Forecasting accuracy evaluation of tourist arrivals," *Ann. Tour. Res.*, vol. 63, pp. 112–127, 2017, doi: 10.1016/j.annals.2017.01.008.
[21] K. Kim, O. joung Park, S. Yun, and H. Yun, "What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management," *Technol. Forecast. Soc. Change*, vol. 123, pp. 362–369, 2017, doi: 10.1016/j.techfore.2017.01.001.
[22] K. Al Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," *Big Data Min. Anal.*, vol. 4, no. 1, pp. 47–55, 2021, doi: 10.26599/BDMA.2020.9020015.
[23] J. Raun *et al.*, "New avenues for second home tourism research using big data: prospects and challenges," *Curr. Issues Tour.*, vol. 26, no. 6, pp. 890–902, 2023, doi: 10.1080/13683500.2022.2138282.
[24] T. Qi, Y. Xu, and H. Ling, "Tourism scene classification based on multi-stage transfer learning model," *Neural Comput. Appl.*, vol. 2, pp. 1–12, 2018, doi: 10.1007/s00521-018-3351-2.
[25] A. Dewangan and R. Chatterjee, "Tourism Recommendation Using Machine Learning Approach," *Adv. Intell. Syst. Comput.*, vol. 564, pp. 447–457, 2019, doi: 10.1007/978-981-13-0224-4.
[26] T. Brandt, J. Bendler, and D. Neumann, "Social media analytics and value creation in urban smart tourism ecosystems," *Inf. Manag.*, vol. 54, no. 6, pp. 703–713, 2017, doi: 10.1016/j.im.2017.01.004.
[27] C. W. Yoo, J. Goo, C. D. Huang, K. Nam, and M. Woo, "Improving travel decision support satisfaction with smart tourism technologies: A framework of tourist elaboration likelihood and self-efficacy," *Technol. Forecast. Soc. Change*, vol. 123, pp. 330–341, 2017, doi: 10.1016/j.techfore.2016.10.071.
[28] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo, "Open innovation and social big data for sustainability: Evidence from the tourism industry," *Sustain.*, vol. 10, no. 9, 2018, doi: 10.3390/su10093215.
[29] Z. Xiang and D. R. Fesenmaier, "Analytics in Smart Tourism Design and Smart Tourism Zheng," *Anal. Smart Tour. Des.*, pp. 299–307, 2017, doi: 10.1007/978-3-319-44263-1.
[30] K. D. Mukhina, S. V. Rakitin, and A. A. Visheratin, "Detection of tourists attraction points using Instagram profiles," *Procedia Comput. Sci.*, vol. 108, no. June, pp. 2378–2382, 2017, doi: 10.1016/j.procs.2017.05.131.
[31] A. Umanets, A. Ferreira, and N. Leite, "GuideMe – A Tourist Guide with a Recommender System and Social Interaction," *Procedia Technol.*, vol. 17, pp. 407–414, 2014, doi: 10.1016/j.protcy.2014.10.248.
[32] V. Subramaniyaswamy, V. Vijayakumar, R. Logesh, and V. Indragandhi, "Intelligent travel recommendation system by mining attributes from community contributed photos," *Procedia Comput. Sci.*, vol. 50, pp. 447–455, 2015, doi: 10.1016/j.procs.2015.04.014.

7% Overall Similarity

Top sources found in the following databases:

- 4% Internet database
- Crossref database
- 0% Submitted Works database
- 4% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	ejournal.utp.ac.id Internet	1%
2	Rania Rizki Arinta, Emanuel Andi W.R.. "Natural Disaster Application on... Crossref	1%
3	coursehero.com Internet	<1%
4	Xin Li, Rob Law, Gang Xie, Shouyang Wang. "Review of tourism forecas... Crossref	<1%
5	researchgate.net Internet	<1%
6	ejournal.uin-suska.ac.id Internet	<1%
7	Filipe Batista e Silva, Mario Alberto Marín Herrera, Konštantín Rosina, ... Crossref	<1%
8	Hamed Taherdoost. "Navigating the Ethical and Privacy Concerns of Bi... Crossref	<1%

9	University of Ulster on 2018-08-07 Submitted works	<1%
10	dspace.kocaeli.edu.tr:8080 Internet	<1%
11	semanticscholar.org Internet	<1%

Excluded from Similarity Report

- Bibliographic material
- Cited material
- Manually excluded sources
- Quoted material
- Small Matches (Less then 10 words)

EXCLUDED SOURCES

journal.irpi.or.id

Internet

67%