

Chris Vassiliadis, Prof.  
University of Macedonia  
Dept. Business Administration  
e-mail: chris@uom.edu.gr  
&  
Nikolaos Stylos, Assoc. Prof.  
School of Management  
University of Bristol, UK  
e-mail: n.stylos@bristol.ac.uk

## **Data Mining**

Data mining (DM) is the knowledge discovery part of database processes (Fayyad, Piatetsky, Shapiro and Smyth, 1996). According to Gartner.com (2021), DM is the process of extracting and discovering meaningful patterns and trends by examining large datasets via sophisticated pattern recognition systems, mathematical and statistical methods. Data sources may comprise databases, data warehouses, information repositories and system data streamed dynamically (Han, Pei and Kamber, 2012).

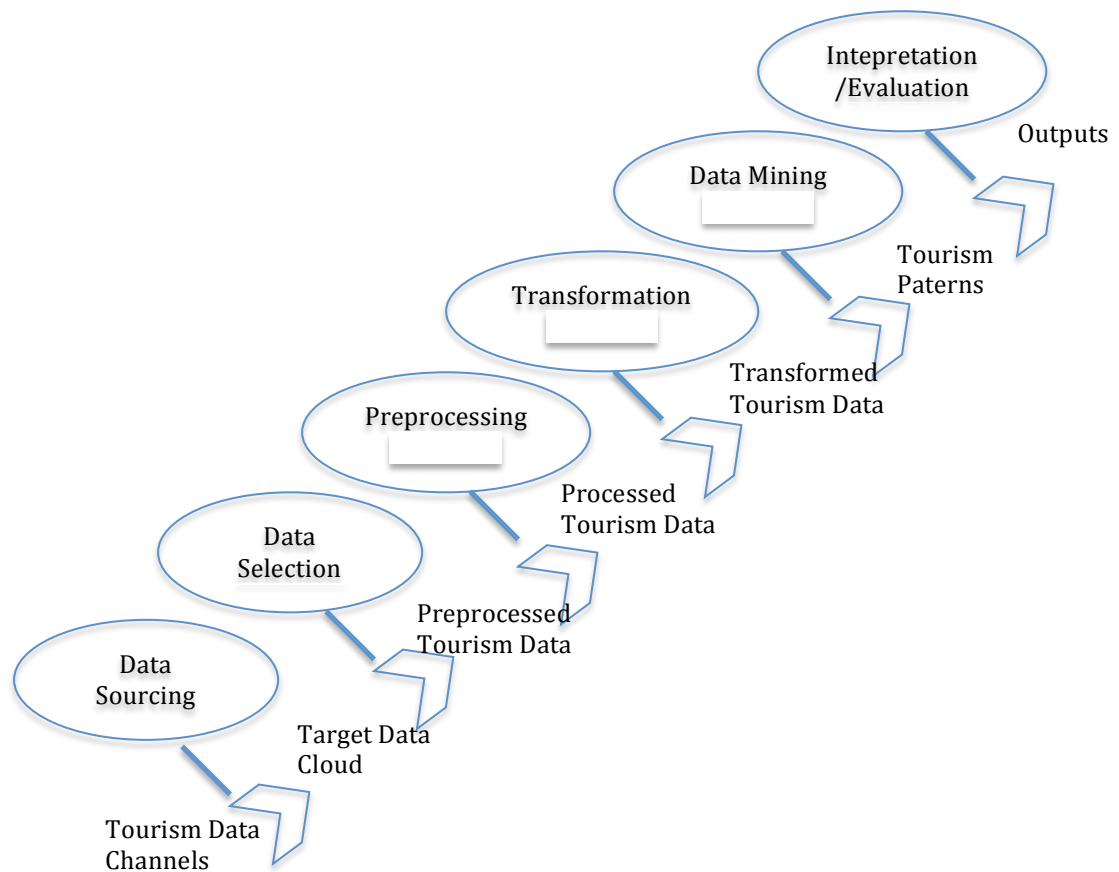
The development of digital hardware components contributed to a DM revolution in the tourism sector beginning in the last two decades of 20th century. At that time, analysts started using non-manual processes for probing datasets with lower data handling costs, lower time handling constraints and lower biases related to subjective human-based judgments (Fayyad *et al.*, 1996). The knowledge of using advanced statistical software in combination with the fast changing advancements and applications of IT and telecommunications in the tourism sector have contributed to building expertise in analysing big datasets to improve tourism stakeholders' decision making (Stylos, Zwiendelaar and Buhalis, 2021). Machine learning and data storage technologies have rapidly changed the forces governing the IT industry and, consequently, the external technological environment of organizations (Stylos, 2019). DM owes its accelerated development to a combination of technological developments and improved practices for importing, processing, analysing, storing and presenting data analysis outputs (see the entry 'Information and Communication Technologies in Tourism' in this Encyclopedia) via a combination of statistical techniques, machine learning, databases, algorithms and information retrieval.

The three main processes comprising DM in tourism are: (1) the discovering of knowledge, patterns and extracting valuable information from tourism-related primary sources; (2) the use of advanced technology that offers automatization in the analysis of large volumes of data; and (3) the support structure for achieving efficiency and effectiveness in tourist decision making (Stylos and Zwiendelaar, 2019).

DM is also increasingly used in applications to identify good practices that contribute to the improvement of tourism services, but also to reduce the cost of operation, production and distribution of these services. It contributes significantly to the determination of customer preferences per customer target

group and can enhance the process of supporting manufacturing and promotional strategies and activities in the conventional and digital environment. It is important for driving efficient processing and data mining in the tourism ecosystem that can be used for future use in the management decision-making process (Agard and Kusiak, 2004; Dam, Dinh and Menvielle, 2019; Karathiya et al., 2012).

As shown in the figure, the key processes that contribute to the gradual conversion of data into useful information in the tourism sector are: (1) sourcing data from multiple information and tourism service exchanges, captured in reports, statistical studies and digital sources; (2) the selection of data from these sources; (3) data clearance, which refers to the preprocessing phase of DM; (4) data homogenization and adaptation for the purpose of measurement and comparison that requires their conversion or transformation; (5) extracting data to obtain appropriate, desirable and qualitatively useful outputs; and (6) interpretation and evaluation for reaching informed decisions via a simple and effective display of the outputs.



**Fig.** The Basic steps of the Data Mining process in the tourism industry

DM is implemented extensively in market-ing studies, financial applications, trading and electronic commerce, security technology, tourism marketing, tourism trajectory and geospace mobility applications (Karathiya et al., 2012; Mazimpaka and Timpf, 2016).

Tourism research and managerial decision making are largely based on analysing multiple datasets arising from large numbers of users sharing/posting textual data, images and videos, via a range of platforms – that is, Wi-Fi, Bluetooth, roaming, RFID data and so on – and functions – web search engine data, booking data, visitor trajectory data and so on as data source creators (Li *et al.*, 2018). DM in tourism has demonstrated that the data extraction process can greatly support the creation of:

- approachable and well-targeted mailing campaigns;
- promotional actions – for example, seasonal campaigns;
- media options and spatial-temporal decisions on ads in advertising campaigns;
- configuration of ad content per recipient of the message;
- determination of the strategies for the selection of profitable business via the formulation of dynamic competitive advantages, linking to certain groups of customers (i.e., B2B and B2C markets);
- support for the room reservation system – for example, in the process of determining the number of rooms for reservation wholesale customers and travellers (i.e., support in the implementation of revenue management).

Support for decision making is facilitated when grouped tourism data storage technologies are introduced alongside other technological advantages such as machine learning, smart and mobile technology, database and artificial intelligence systems. In this line, Xiao and Xiang (2017) postulated that data association analysis, sequence patterns, data classification processes and data storage technologies – for example, using cloud technologies – can significantly contribute to tourism-related managerial decision support systems.

The penetration of various digital platforms in the lives of tourists has introduced changes in managing data, making DM crucial for processing and extracting knowledge with respect to the level of tourist satisfaction, product characteristics and attractiveness, tourist destinations and respective travel routes (Li *et al.*, 2018). DM can prove invaluable in addressing limitations in analysing and forecasting tourist flows and expenditures, as well as conducting market segmentations; taking advantage of the scattered and large volumes of data available is a necessary condition for the effectiveness of the overall tourism destination decision support process.

Tourism destination managers and place marketing managers are interested in improving the place image, and in introducing appropriate tourism services and facilities for supporting the tourism development and planning (Buhalis, López and Martínez-Gonzalez, 2020). The use of targeted data that could support decision making on these issues is now crucial to take advantage of the scattered

and large volume of tourism-related data and a necessary condition for the effectiveness of tourism destination decision support processing systems (Miah *et al.*, 2017).

## References

Agard, B. and Kusiak, A. (2004), 'Data-mining-based methodology for the design of product families', *International Journal of Production Research*, 42(15), 2955–69.

Buhalis, D., López, E.P. and Martínez-González, J.A. (2020), 'Influence of young consumers' external and internal variables on their e-loyalty to tourism sites', *Journal of Destination Marketing & Management*, 15, Article 100409.

Dam, N.A.K., Dinh, T.L. and Menvielle, W. (2019), 'Marketing intelligence from data mining perspective – a literature review', *International Journal of Innovation, Management and Technology*, 10(5), 184–90.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), 'From data mining to knowledge discovery in databases', *AI Magazine*, 17(3), 37–54.

Gartner.com (2021), 'Data mining', accessed 25 March 2021 at <https://www.gartner.com/en/information-technology/glossary/data-mining>.

Han, J., Pei, J. and Kamber, M. (2012), *Data Mining: Concepts and Techniques* (3rd edition), Burlington, MA: Morgan Kaufmann.

Karathiya, M.B., Sakshi, S.R., Sakshi, D.S. and Karathiya, D.R. (2012), 'Data mining for travels and tourism', *Journal of Information and Operations Management*, 3(1), 114–18.

Li, J., Xu, L. and Tang, L. *et al.* (2018), 'Big data in tourism research: a literature review', *Tourism Management*, 68, 301–23.

Mazimpaka, J.D. and Timpf, S. (2016), 'Trajectory data mining: a review of methods and -applications', *Journal of Spatial Information Science*, 13, 61–99.

Miah, S.J., Vu, H.Q., Gammack, J. and McGrath, M. (2017), 'A big data analytics method for tourist behaviour analysis', *Information & Management*, 54(6), 771–85.

Stylos, N. (2019), 'Technological evolution and tourist decision-making: a perspective article', *Tourism Review*, 75(1), 273–8.

Stylos, N. and Zwiendelaar, J. (2019), 'Big data as a game changer: how does it shape business intelligence within a tourism and hospitality industry context?', in M. Sigala, R. Rahimi and M. Thelwall (eds), *Big Data and Innovation in Tourism, Travel, and Hospitality*, Singapore: Springer, pp. 163–81.

Stylos, N., Zwiigelaar, J. and Buhalis, D. (2021), 'Big data empowered agility for dynamic, volatile, and time-sensitive service industries: the case of tourism sector', *International Journal of Contemporary Hospitality Management*, 33(3), 1015-36.

Xiao, W. and Xiang, C. (2017), 'Overview of tourism data mining in big data environment', in *Proceedings of the 2016 7th International Conference on Education, Management, Computer and Medicine (EMCM 2016)*, Dordrecht: Atlantis Press.