Data Mining Techniques and Its Application in Civil Engineering—A Review

Priyanka Singh

1 Introduction

Data mining is the technique of discovering styles of massive data units regarding strategies at the intersection system mastering, facts and database systems. It is the workout of analyzing massive pre-cutting-edge databases in order to generate new records. Data mining is the assessment step of the “information discovery in database” [1, 2]. It is an interdisciplinary subfield of computer science and statics with a preferred aim to extract facts from records set and rework the facts into an understandable shape for further use. This method is likewise used in corporation programs along with market segmentation, fraud detection, and credit danger evaluation further to many different programs [3]. It has the capacity to locate pattern saved inner records and is now taken into consideration as a catalyst for growing business enterprise procedure by using keeping off failure sample. There is a frequent collection of data on each day basis in construction and manufacturing companies while operations. The main intention of data mining is to extract data from the pre-existing dataset for future use in the logical structure. It is now considered as a catalyst for increasing business process by avoiding failure pattern [4, 5].

1.1 Methods and Approach

Some of the applications which are being utilized in the field of civil engineering are as follows [6–9]:

P. Singh (✉)
Department of Civil Engineering, Amity School of Engineering & Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: priyanka24978@gmail.com

P. K. Kapur et al. (eds.), *Advances in Interdisciplinary Research in Engineering and Business Management*, Asset Analytics,
https://doi.org/10.1007/978-981-16-0037-1_15

i. Predicting strength based on the concrete mixture, soil, and several other variables.
ii. Monitoring infrastructure health of sub- and super-structure on the data received from the sensors.
iii. Traffic Engineering: Data sensing, analysis, and mining can be used to facilitate decision-making and intelligent transportation systems.
iv. Spatial data mining to find out the best possible location for construction.
v. Water resources engineering: In water resources engineering, data mining is used to identify chemical trends in water quality samples.
vi. Construction Management: It is used to estimate maintenance cost and quality of the construction.
vii. Environmental Engineering: It could be used to study environmental and natural resource science.
viii. Highway and Transportation Engineering: Traffic and pavement management can be done with the application of data mining in the field of highway and transportation engineering.
ix. Hydraulics and Water Power Engineering: The prediction of dangerous risks in Hydraulics and Water Power Engineering can be achieved by data mining techniques.
x. Materials Science and Engineering: Data mining techniques can be applied in many domains such as astronomy, bioinformatics, chemistry, materials science, climate, fusion, and combustion.
xi. Geotechnical Engineering: Data mining can be explored to formulate several complex geotechnical engineering problems. The complex analysis of geotechnical behavior is due to multivariable soil and rock responses.
xii. Earthquake Engineering: Application of data mining would be extremely helpful in the prediction of natural geological calamities like flood, tornado, hurricane, volcanic eruption, earthquake, heat-wave, or landslides.
xiii. Coastal and Harbor Engineering: By data mining, the different wave behaviors at the coastal areas can be studied, which will be useful for designing harbor engineering.
xiv. Tunnel Engineering: The variation in the pattern of geological data, stress–strain data of supporting structures and the deformation data of the surrounding rocks can be interpreted by the data mining technique.
xv. Surveying and Geo-Spatial Engineering: The use of new and developing technologies such as GPS, satellite imagery, laser mapping, and fast computing to create complex layers of interconnected geographic information can be facilitated by data mining technique.
xvi. Geomatics: Application of data mining is utilized in delivering spatially referenced information by the data of deposit collection, processing, interpreting, storing of mining, and extraction plant.
xvii. Geosciences: By the application of data mining, petro-physical data, logging data, seismic data, and geological data can be addressed.
xviii. Remote Sensing: Data mining is utilized to extract aerial remote sensing imagery for automatic land-cover classification.

xix. Geographical Information systems: Data mining can automatically extract knowledge from raw data of Geographical Information Systems to gain spatial analysis result.

The field of construction as well as other industries related to it makes the big statistics technology utilize the large quantity of information received and saved through advanced computing systems. Information comes from anywhere: computers, structures, human beings, sensors, and any device that generates records. In the construction field, every structure that has been erected embodies huge quantities of data [6, 10, 11]. However, managers and civil engineers need to explore the answers hidden in unstructured information. Big data technology gives civil engineers the strength to utilize unstructured information, which is tough to accumulate and analyze manually in any beneficial manner. In fact, all records are useless without correct assessment and verification [10, 12].

2 Application of Data Mining Technology on Construction Project Cost Control System

With the development of construction management, it is essential that the costing-related tasks are done correctly. The innovation regarding data mining process in the development of the cost control framework to resolve the deficiencies in conventional management is highly beneficial [13–15].

Secondarily, the main focus is on direct expenses on items rather than on overhead cost by cost management programs. Rational cost and cost of the quantity of items are often neglected because much more importance is given to financing [16]. There is great redundancy of data in the information system thus making it different to determine variation in the information. Usually issues like selection of important data from huge arise. Hence Data Mining Technology comes as a solution to the problem [17].

The process of sorting the important data from large data storage and extracting relatable patterns to solve problems with ease is facilitated with DM [18]. The user must have adequate knowledge to discover such databases. As a result, Data Mining Technology grows with great importance in the field of multiple costing parameters.

Data Mining Technologies plays a great role in Cost Management. Several goals that are achieved are as follows:-

- Rapid or quick schedule
- Reduction in cost
- Availability of good quality data values
- On-time targets.

Manual recording of data in a common system did not have the provision of accessibility which lead to the problem of missing data. As a result, data were classified into three dimensions:

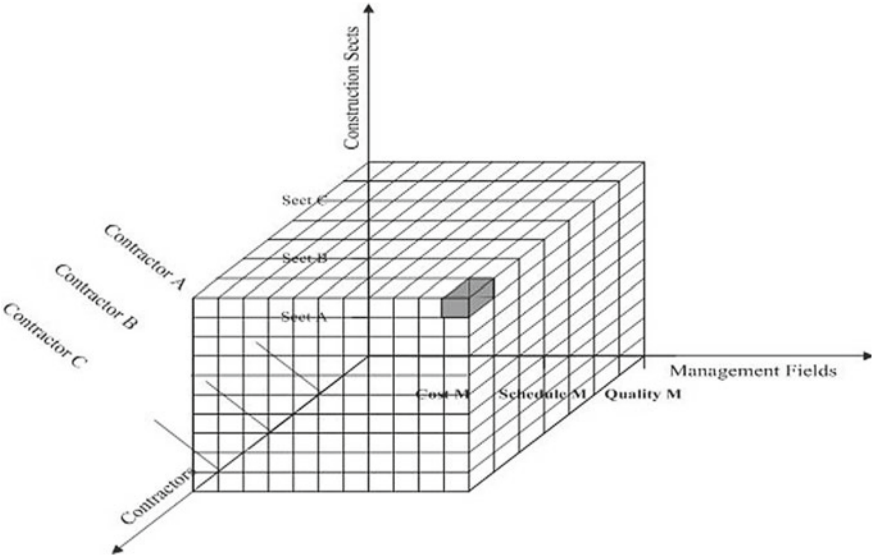

Fig. 1 Three-dimensional data model [20]

- Management field
- Contractors field
- Construction sector.

Initially the data were in one dimension. Now 3D projection and storage of data information has overcome the issue of missing data and has made it easier [19]. Figure 1 illustrates the three-dimensional data model of data information.

The Management field in 3D model can further be separated into three sections (i) cost management, (ii) schedule management, and (iii) quality management. Figure 2 depicts the picture of data dividing in three-dimensional model. Hence special information might be consolidated into a new datum [21].

3 Application of Big Data in Construction Industry

The application of big data in the construction Industry consists of basic three processes:

- Design: Big data analyzes the building design and its modeling. It can also be used to finalize where the building can be constructed and with which material. However, old records can be studied with it to know the probability of risks at the construction site.
- Build: During construction at the site, the data collected from weather, traffic, and other activity is interpreted to decide the stages of construction.

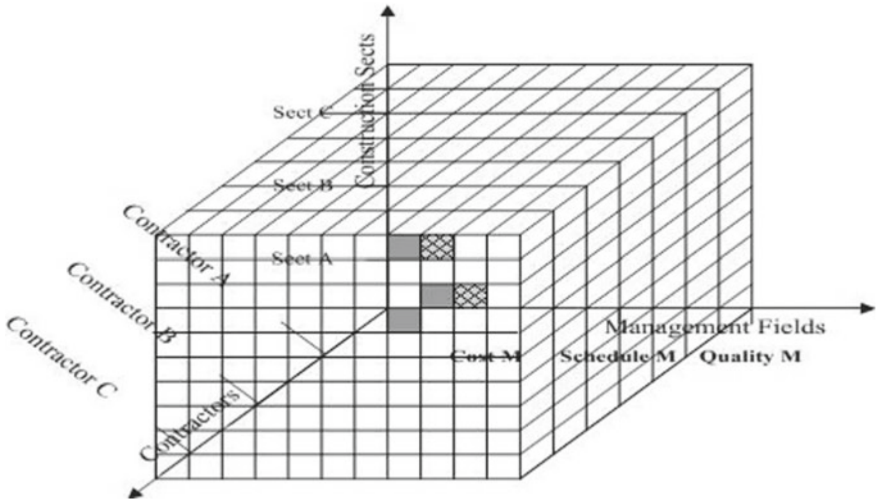

Fig. 2 Data dividing in three-dimensional model [22]

- Operate: The accomplishment of the construction at each level can be tracked by collecting the data from the sensors built into the super and substructures of the buildings. Scheduled maintenance activity can be planned by feeding the data of the sensors in the building information modeling.

The handling of data gets bigger as the size of the data increases. This aspect can be managed by big data analytics. Big data analytics provide inspection alert much before the risk levels [23, 24].

There is a frequent collection of data on daily basis in construction companies for activities and operations. The job overheads increases proportionally with the number of changed orders during the month for high productivity. Projects hidden within the project database can be detected using data mining [25]. The flow chart of the data mining process is shown in Fig. 3.

Data mining serves two objectives, namely Insight and Prediction. Insight prompts in recognizing design and patterns. Prediction gives expectations dependent on information [27]. Large data mining techniques of discrete types, for example, neural networks and conceptual clustering are used in particular domains individually. Data sensing, analysis, and mining can be used to facilitate decision-making and intelligent transportation systems [28].

4 Data Mining in the Field of Transportation Engineering

The study of data mining in the field of Transportation has a great scope. Various problems were acknowledged through data mining such as.

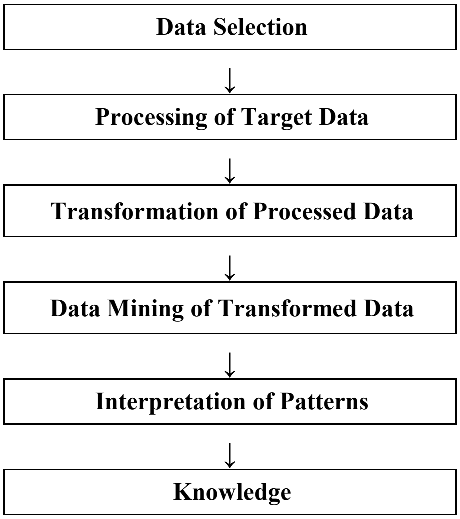

Fig. 3 Flow chart of data mining process [26]

- Productive methodology for dealing with large information with accurate results
- Methods to differ over-fitting
- To check absence and uproarious information
- Build connections between distinct fields
- Better client communication and pre-learning.

Specialized forms of data help to monitor cases like drowsy drivers by sensing their state of driving. It also helps in designing roadways keeping in mind the overtaking speed of vehicles and the rate of accidents [29]. Table 1 illustrates Data Characteristics in GIS Application of transportation engineering.

Table 1 Data characteristics in GIS application of transportation engineering [30]

	Logical	Physical
Real World	*Legal definitions*	*Actual facilities*
	Route	Highways
	State Trunk Network	Roads
	Country Trunk Network	Interchanges
	Street network	Intersections
	Political Boundary	
Virtual World	*Data structures*	*Data values*
	Networks	Lines
	Chains	Points
	Links	Poly-lines
	Nodes	Polygons
	Lattices	Attributes

Huge amounts of data are given by PMS for pavement design, the number of vehicles running on that particular pavement, types of materials used to construct that pavement, undulations present, etc. Data provided by GPS helped in tracing the vehicles and to check the inconvenient changing of lanes. In geographic data frameworks (GIS) for transportation, interconnected equipment, programming, information, individuals, and assessment of data are done, which is to be considered as a vital job [31–34].

5 Data Mining in the Field of Environmental Engineering

In the field of environmental engineering, the aspect of data mining can be utilized to the optimum. Data mining techniques in environmental studies provide valuable knowledge and useful patterns which facilitate us to deal with environmental problems, like air pollution, water pollution, etc. [35, 36]. Some of the beneficial aspects of data mining in environmental engineering are as follows:

- Prediction of rainfall based on the parameters of environmental studies.
- Prediction of rainfall and climate change can be predicted on the different patterns obtained by the data mining techniques. [37, 38]
- Ecological systems can be modeled by processing ecological and environmental data [39, 40]
- Better control of air quality and water quality by analyzing environmental data.
- Better waste management and recycling can be accomplished by discovering patterns through data mining [38, 41].

6 Conclusion

This paper shows the various application related to civil engineering where big data plays a very big role in the management of project, maintaining resources, scheduling of jobs, estimation of time and cost involved for civil construction works. Big data analytics along with data mining can make the construction industry fruitful and suggests the methods for improving the health of construction. Data available at various construction sites may be gathered to form big data and the data mining methods like clustering and classification algorithms can be implemented for better analysis, decision–making, and prediction of the new estimates using the past data. Data mining has huge scope in the field of environmental engineering also. It facilitates to have better control over air and water quality. Prediction of the weather forecast and climate change can be optimized.

References

1. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining.* Cambridge, MA: AAAI Press/The MIT Press.
2. Han, J. &, Kamber, M. (2001). *Data mining: Concepts and techniques.* Higher Education Press.
3. Hall, M. J., Minns, A. W., & Ashrafuzzaman, A. K. M. (2002). The application of data mining techniques for the regionalisation of hydrological variables. *Hydrology and Earth System Sciences, 6,* 685–694.
4. Chau, K. W., & Cao, Y. (2002). The Application of data warehouse and decision support system in construction management. *Automation in Construction, 12,* 213–224.
5. Zhang, J. P., & Wang, H.-J. (2002). Towards 4D management for construction planning and resource utilization. In *The 9 th International Conference on Computing in Civil and Building Engineering*, Taiwan (pp. 1281–1286).
6. Hyperion Software Corp. (1999). The role of OLAP server in a data warehousing solution.
7. Chakrabarti, S. (2002), Mining the web: Statistical analysis of hypertex and semi-structured data. Morgan Kaufmann.
8. . Hong-Yan, L. I, Bu-Ying, C., & Li, D. (2013). Application of data warehouse and data mining in coal information management (Vol. 31, no. 8, pp. 31–32).
9. Forbes,L. H., & Ahmed, S. M. (2003). Construction integration and innovation through lean methods and E-business applications, construction research 2003, Copyright ASCE 2004.
10. Inmon, W. H. (2000). *Building the data warehouse* (2nd ed.). China Machine Press.
11. Bilal M., & Oyedele, O. L. (2016). Big Data Architecture for Construction Waste analytics (CWA): A conceptual framework. *Journal of Building Engineering*, 144–156.
12. Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. Wiley.
13. Hore, A. (2006). Use of IT in managing information and data on construction projects—A perspective for the IRISH construction industry, information technology in construction project management.
14. Zhou, Y., & Ding, L. Y. (2006) International symposium on “Advancement of Construction Management and Real Estate” The CRIOCM 2006.
15. Jin, C. (2017). Real-time damage detection for civil structures using Big Data.
16. Adrians, P., & Zantinge, D. (1996). *Data mining*. England: Addison-Wesley Longman.
17. Cabena, P. (1997). *Discovering data mining: From concept to implementation.* NJ: Prentice Hall.
18. Han, J. (2001). *Data mining: Concepts and techniques.* San Francisco: Morgan Kaufmann Publishers.
19. Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Massachusetts: MIT press.
20. Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). Morgan Kaufmann.
21. Attoh-Okine, N. O. (1997). Rough set application to data-mining principles in pavement management database. *Journal of Computing in Civil Engineering, American Society of Civil Engineers, 11*(4), 231–237.
22. Soibelman, L., & Hyunjoo, K. (2002). Data preparation process for construction knowledge generation through knowledge discovery in databases. *Journal of Computing in Civil Engineering, ASCE, 16*(1), 39–47.
23. Jae-Gil Lee, M. K. (2015), Geospatial Big Data: Challenges and opportunities. *Big Data Research*, 74–81.
24. Ahmad, I., & Ahmed, S. M. (2001). Integration in the construction industry: Information technology as the driving force. In R. L. K. Tiong (Ed.), *Proceedings of the 3rd International Conference on Construction Project Management*. Nan yang Technical University Press.
25. Hu, D. (2005). Research on the systematic framework of computer integrated construction. Huazhong University of Science and Technology.
26. Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques.* Higher Education Press.

27. Rezania, M., Javadi, A., Giustolisi, O. (2008). An evolutionary-based data mining technique for assessment of civil engineering systems.
28. Shanti, M. A., & Saravanan, K. (2017). Knowledge data map—A framework for the field of data mining and knowledge discovery. *International Journal of Computer Engineering & Technology, 8*(5), 67–77.
29. Barai, S. V., & Reich. (2001). Data mining of experimental data: Neural networks approach. In *Proceedings of 2nd International Conference on Theoretical, Applied Computational and Experimental Mechanics ICTACEM (CD-ROM).*
30. Rabee M. Reffat, John S. Gero, Wei Peng (2004). Using data mining on building maintenance during the building life cycle.
31. Kohavi, R. (2001). Data mining and visualization. In *Sixth Annual Symposium on Frontiers of Engineering* (p.p. 30--40). National Academy Press, D. C.
32. Amado, V. (2000). Expanding the use of pavement management data. In *Transportation Scholars Conference*, University of Missouri.
33. Tan, P., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*. Addison Wesley.
34. Dzeroski, S. (2003). Environmental applications of data mining. Lecture Notes of Knowledge Technologies, University of Trento.
35. Stojic, A., Stojic, S. S., Reljin, I., Cabarkapa, M., Sostaric, A., Perisic, M., & Mijic, Z. (2016). Comprehensive analysis of PM10 in Belgrade urban area on the basis of long-term measurements. *Environmental Science and Pollution Research, 23*, 10722–10732. https://doi.org/10.1007/s11356-016-6266-4.
36. Gaal, M., Moriondo, M., & Bindi, M. (2012). Modelling the impact of climate change on the Hungarian wine regions using random forest. *Applied Ecology and Environmental Research, 10,* 121–140. https://doi.org/10.15666/aeer/1002_121140.
37. Crimmins, S. M., Dobrowski, S. Z., & Mynsberge, A. R. (2013). Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling, 266,* 126–130. https://doi.org/10.1016/j.ecolmodel.07.006.
38. Lei, K. S., Wan, F. (2012). Applying ensemble learning techniques to ANFIS for air pollution index prediction in Macau. In *International Symposium on Neural Networks (ISNN'12)*, 11–14 July 2012 (pp. 509–516). Berlin, Heidelberg: Springer.
39. Budka, M., Gabrys, B., & Ravagnan, E. (2010). Robust predictive modelling of water pollution using biomarker data. *Water Research, 44,* 3294–3308. https://doi.org/10.1016/j.watres.2010.03.006.
40. Singh, K. P., Gupta, S., & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environment, 80,* 426–437. https://doi.org/10.1016/j.atmosenv.2013.08.023.
41. Nelson, T. A., Coops, N. C., Wulder, M. A., Perez, L., Fitterer, J., Powers, R., & Fontana, F. (2014). Predicting climate change impacts to the Canadian Boreal forest. *Diversity, 6,* 133–157. https://doi.org/10.3390/d6010133.