

# SURVEY OF DATA MINING ALGORITHMS FOR INTELLIGENT COMPUTING SYSTEM

S. Iwin Thanakumar Joseph

Department of Computer Science and Engineering,

Karunya Institute of Technology and Sciences,

Coimbatore, India.

[iwineee2006@gmail.com](mailto:iwineee2006@gmail.com)

**Abstract:** The Intelligent computing system, described to be a collection of the connected device working in mutual understanding to attain a particular purpose, is an incorporation of artificial intelligence and the computational intelligence, and are employed in variety of applications. The paper presents the survey on the data mining algorithms and the techniques that could be employed with the intelligent computing system, presenting a basic conception of the data mining along with the prominent algorithms of the data mining and the classification of its techniques, further the survey concludes with the challenges included in the overview of the survey done along with the future enhancement in the research that analyses the data mining techniques in the intelligent computing applications.

**Keywords:** Intelligent Computing System, Data Mining Algorithm's, Artificial Intelligence, Computational Intelligence and Data Mining Techniques

## 1. Introduction

The capacity of the artificial intelligence and its progressive growth had paved for its incorporation in multitudes of intelligent methodologies, to develop intelligent computing systems. The intelligent computing systems would enables to make proper decisions even with the complex problems, the intelligent system integrating the expert systems has concentered a considerable amount of achievements. But the experts systems faces difficulties in the acquisition and the processing of the knowledge. The knowledge awareness or the expertise from each field and the variations in them causes challenges and issues when dealing with the intelligent computing system that are used along with the expert system. [10] , so in order to recognize the pattern involved and the gain the knowledge from various fields it becomes essential to involve the data mining in the intelligent computing system to gather the pattern involved and the information present in it. The fig.1 Shows the intelligent computing systems frame work.

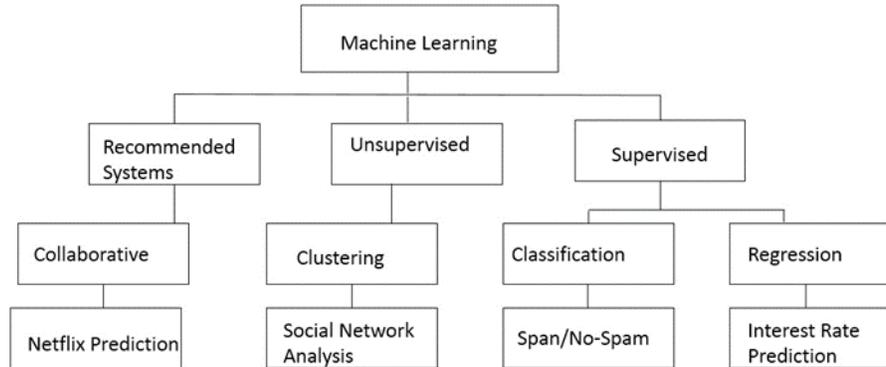


Fig. 1 Intelligent Computing System

The Data Mining is the multi-disciplinary subfield of the computer science technology and is primitively a process of computing for discovering patterns of big data sets, it is an indispensable process including the intelligent methods in the for retrieving the patterns of data. The data mining are termed as the practice of analysing enormous prevailing dataset for generation of new information, otherwise known as the process of knowledge discovery from the data base. In most of the applications the data or the knowledge gained from the data mining are optimistically new and very profitable, the data or the information extracted using the date mining are usually reserved for the later use. It employs one or more software's for examining of big sets of data to extract the patterns hidden in to it. In a this world filled with business tactics everywhere and every minute, the data mining techniques and the methodologies evolve as the efficient methodologies of analyzing the raw data and converting them into insight that help in making the important business decisions. To be explicit the data mining techniques dig in deep utilizing the materials, in order to help us understand the aspects of the various elements. For e.g.: the huge amount of information's gathered from the internet giving the details of various fields could be referred as the collection of raw data containing details in it. The same data becomes a knowledge or insight leading to a certain development if sorted out properly, and this could be done using the data mining.

So the paper is to present the survey on the data mining techniques in the intelligent computing systems, presenting the basics of the data mining along with its available methods and the algorithms and their efficiency in the intelligent computing system.

The remaining paper is organized with the methods of data mining with the basic information related to the data mining in section 2 and the literature survey of the data mining algorithms for the intelligent computing systems in section 3 and the conclusion in section 4

## 2. Data Mining Methods

Data mining defined to be the method that extracts, usable insights from the set of larger data sets, of collection of raw information, entails the analyzing of the patterns hidden in the information. So it is used in the wider applications of the belonging to different areas like, science, research, business, industries etc. So data mining can be viewed as the optimum method of knowledge discovery and warehousing along with the computing, to provide with better insight leading to improved decision making [2]. Some of the key features that are involved in data mining methods and its technical infrastructure requirement are [13] listed below

### (a) Key Features

- (1). Flow and attitude analysis by pattern recognition,
- (2) Development of decision-adapted information,
- (3) Prediction from the similar outcomes,
- (4) Analysis of large data sets, and
- (5) Detection of hidden facts using the clustering

### (b) Further its technical infrastructure requires

- (i) Volume of the Data Size: A huge amount of data are to be processed and maintained, in order to gain an insight paving a more useful and powerful systems.
- (ii) Powerful Systems: The processing or computing of the more complex and the greater number of queries requires a powerful system for handling them more efficaciously. The fig.2 below shows the implementation process involved in the data mining.



Fig .2 Process of Data Mining

The process of the data mining include the collection of raw data belonging to data bases such as the relational, data warehouse, advanced and information reserves, object-oriented and object-relational, transaction and spatial, heterogeneous and legacy, multimedia and streaming, text, text mining and web mining. The process involved in the data mining to derive insights makes them to be recognized with various names such as knowledge discovery, information harvesting, pattern analysis and knowledge extraction etc. enables us to understand the usefulness of the data, paving to a constructive measures of the area involved. The fig .3 given below lists the data mining methods in retrieving the useful information's from the huge sets of data gathered.

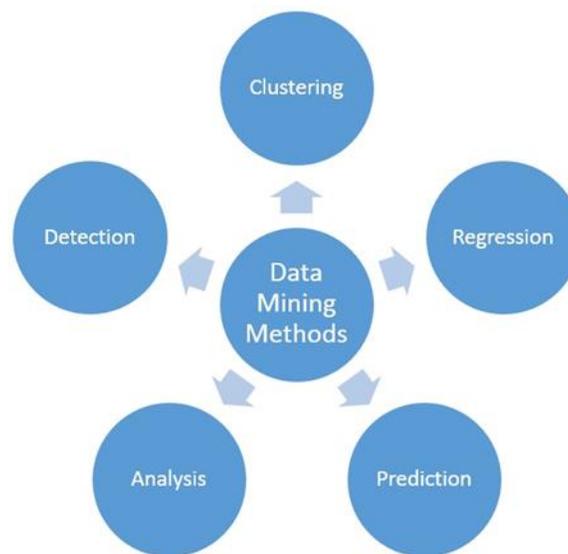


Fig .3 Data Mining Methods

Each method of data mining enables us to identify the hidden aspects of the various fields and their applications. Where the data mining method of clustering enables us, to have the insights of the differences and the similarities from the data gathered, in which the sequential patterns allow us to have the similarities identified in the patterns or the trends [17], [18], the analysis enables one to collect the more relevant and the vital information's from the data and the meta data gathered. [20], the detection allows us to gain an insight on the dissimilarities in the expected attitude or the patterns, and could be employed in various application such as intrusion, fault [22], fraud [23] and disease detection [12] the regression gives the relationship between the variables in order to identify the possibility of the

specific variable with the other variables, and finally the prediction involves all the above mentioned methods and analyses the present and the past data and predicts the future happenings [15], [16]

### 3. Data Mining Algorithms in Intelligent Computer System [7] [20] [12][14][17]

The survey gathered based on the data mining algorithms in the intelligent computing system the section below provides the in depth description of the various algorithms of the data mining, some of the prominent algorithms of the data mining.

Clustering algorithm: This clustering algorithm for the data mining is the segmentation algorithm that frames clusters based on the similarities in the data. The clustering algorithms takes multiple strides (iterations), to frame the clusters of the data with similar attributes. The data mining utilizing the clustering algorithm is capable of generating an analysis without citing the predictable columns, the usually proceed initially with the identification of the relation in the database gathered and produces a sequential classification, and each instance in a dataset could be represented using the scatter diagrams. The fig.4 shows the scatter diagram.

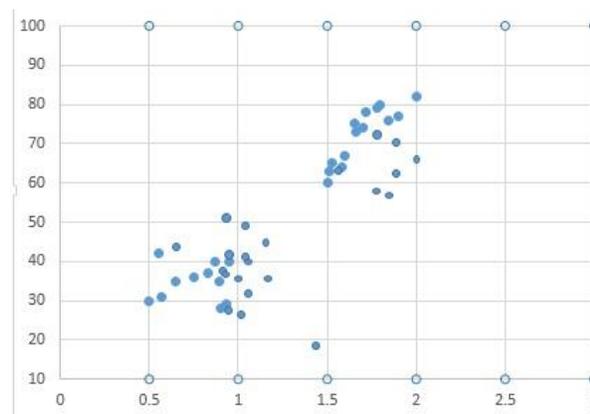


Fig .4 Scatter Diagrams

Neural Network Algorithm: The neural network algorithm based on the MLP is equipped with the three layers of perceptron's or the neurons, which are the input, output and the hidden layer, and is also known as the back propagation delta rule network. In the neural network the information provided in the input is produced at the output as the simple non-linear function that is the summation of the input. The nodes of the three layers are connected to each other, but

the nodes present in the particular layer are not interconnected. The input provided is connected to the nodes in the hidden layer and then passed to the output layer. Neural network for the data mining can be employed with the multiple networks for the purpose of prediction. The input layer engaged with the inputs given, presents the input properties and based on its values either as false or true and based on its existence as missing or existing. The hidden layer accepts the input form the input layer and sends it to the output layer, the output layer produces the predicting model for the data mining, presenting the future prediction of the input events. The fig.5 gives the frame work of the neural network.

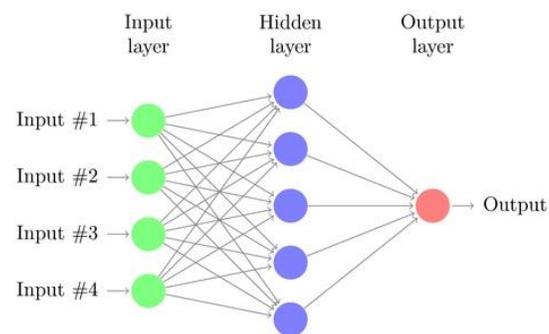


Fig .5 Neural Network Frame Work

Regression Algorithm: The linear regression algorithm in data mining presents the relationship between the data observed, and frame the sequence of the information's observed.

Bayesian Algorithm: The predictive modelling of the data mining can also done utilizing the Bayesian algorithm, this algorithm involves the less number of computations compared to the other algorithms as the algorithm determines the conditional probabilities of the input and the output observed, and performs a quick analysis presenting the relationship between the input and the output. This algorithm is prominent in the initial stage detection and could be used in the diagnosis of the infections or harmful disease for the early detection.

### 3.1. Challenges and the Benefits Incurred in Data Mining

Despite of promising nature of the data mining, in the hidden insights of data procured, the also incur the following challenges list below

- Requires experts for the formulation of the data mining.
- The structure of the data mining might be over fitting as they limited size of the database used for training would befit the future declarations.
- Requires a very large set of data sets that is sometimes presents hazards in handling them, in terms of storage, computation etc.
- Requires alteration in the regular practices undertaken for recovering the deep insights for e.g.: (Business, industries etc.)
- The accuracy of the data mining results obtained depends on the diversity of the data set gathered.
- The combination of different types of information on the case of the heterogeneous data set and the global information system might be complicated

The following presents the benefits of the Data Mining

- Provides the knowledge base on the specific area
- Improves the profit rate of the organization
- Enhances the decision making
- Automation in the prediction and discovery
- Can be implemented with the novel system at ease
- Ensures a high speed powerful process for one to examine the a huge amount of information's

### 4. Conclusion

The paper presenting the survey of the data mining algorithms in the intelligent computing, system s presents with the, basics of the data mining in the initial stage followed by the methods that could be utilized for data mining and later proceeds with the some of the prominent algorithms of the data mining indulged in the intelligent computing systems and finally presents challenges and the benefits of utilizing the data mining. The data mining could be used in multitude of applications in the various fields not limited to the healthcare, telecom, bank sectors , insurance,

manufacturing, education, service provisioning, retail marketing, bio-informatics, etc. further as the future enhancement the paper is to continue with the implementation of the data mining in the application of the intelligent computing.

## References

- [1] Yang, Qiang, and Xindong Wu. "10 challenging problems in data mining research." *International Journal of Information Technology & Decision Making* 5, no. 04 (2006): 597-604.
- [2] Maimon, Oded, and Lior Rokach, eds. "Data mining and knowledge discovery handbook." (2005).
- [3] Tsihrintzis, George A., Maria Virvou, and Lakhmi C. Jain, eds. *Intelligent Computing Systems: Emerging Application Areas*. Vol. 627. Springer, 2016.
- [4] Najafabadi, Maryam Khanian, Azlinah Hj Mohamed, and Mohd Naz'ri Mahrin. "A survey on data mining techniques in recommender systems." *Soft Computing* 23, no. 2 (2019): 627-654.
- [5] Maji, Srabanti, and Srishti Arora. "Decision Tree Algorithms for Prediction of Heart Disease." In *Information and Communication Technology for Competitive Strategies*, pp. 447-454. Springer, Singapore, 2019.
- [6] Aljarah, Ibrahim, Majdi Mafarja, Ali Asghar Heidari, Hossam Faris, and Seyedali Mirjalili. "Multi-verse optimizer: theory, literature review, and application in data clustering." In *Nature-Inspired Optimizers*, pp. 123-141. Springer, Cham, 2020.
- [7] Zhang, Junlin, Samuel Oluwarotimi Williams, and Haoxiang Wang. "Intelligent computing system based on pattern recognition and data mining algorithms." *Sustainable Computing: Informatics and Systems* 20 (2018): 192-202.
- [8] Iqbal, Rahat, Faiyaz Doctor, Brian More, Shahid Mahmud, and Usman Yousuf. "Big data analytics: Computational intelligence techniques and application areas." *Technological Forecasting and Social Change* (2018).

- [9] Duan, Yanqing, John S. Edwards, and Yogesh K. Dwivedi. "Artificial intelligence for decision making in the era of Big Data—evolution, challenges and research agenda." *International Journal of Information Management* 48 (2019): 63-71.
- [10] Yellamma, Pachipala, N. arasimham Challa, and V. Sreenivas. "Intelligent Data Security in Cloud Computing." *International Journal of Current Engineering and Technology* 4, no. 1 (2014).
- [11] Tan, Pang-Ning. *Introduction to data mining*. Pearson Education India, 2018.
- [12] Souri, Alireza, and Rahil Hosseini. "A state-of-the-art survey of malware detection approaches using data mining techniques." *Human-centric Computing and Information Sciences* 8, no. 1 (2018): 3.
- [13] <https://economictimes.indiatimes.com/definition/data-mining?from=mdr>
- [14] Paul, Seth, Jamie MacLennan, Zhaohui Tang, and Scott Oveson. "Data Mining Tutorial." *Microsoft Corporation* (2005): 50-59.
- [15] Faradonbeh, Roohollah Shirani, and Abbas Taheri. "Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques." *Engineering with Computers* 35, no. 2 (2019): 659-675.
- [16] Goswami, Saptarsi, Sanjay Chakraborty, Sanhita Ghosh, Amlan Chakrabarti, and Basabi Chakraborty. "A review on application of data mining techniques to combat natural disasters." *Ain Shams Engineering Journal* 9, no. 3 (2018): 365-378.
- [17] Berkhin, Pavel. "A survey of clustering data mining techniques." In *Grouping multidimensional data*, pp. 25-71. Springer, Berlin, Heidelberg, 2006.
- [18] Wang, Ruili, Wanting Ji, Mingzhe Liu, Xun Wang, Jian Weng, Song Deng, Suying Gao, and Chang-an Yuan. "Review on mining data from multiple data sources." *Pattern Recognition Letters* 109 (2018): 120-128.

- [19] Fernandes, Eduardo, Maristela Holanda, Marcio Victorino, Vinicius Borges, Rommel Carvalho, and Gustavo Van Erven. "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil." *Journal of Business Research* 94 (2019): 335-343.
- [20] Bai, BG Mamatha, B. M. Nalini, and Jharna Majumdar. "Analysis and Detection of Diabetes Using Data Mining Techniques—A Big Data Application in Health Care." In *Emerging Research in Computing, Information, Communication and Applications*, pp. 443-455. Springer, Singapore, 2019.
- [21] Gaber, Mohamed Medhat, Adel Aneiba, Shadi Basurra, Oliver Batty, Ahmed M. Elmisery, Yevgeniya Kovalchuk, and Muhammad Habib Ur Rehman. "Internet of Things and data mining: From applications to techniques and systems." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9, no. 3 (2019): e1292.
- [22] Yang, Hsu-Hao, Mei-Ling Huang, Chun-Mei Lai, and Jhih-Rong Jin. "An approach combining data mining and control charts-based model for fault detection in wind turbines." *Renewable energy* 115 (2018): 808-816.
- [23] Al-Radaideh, Qasem A., and Mahmoud M. Al-Zoubi. "A data mining based model for detection of fraudulent behaviour in water consumption." In *2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 48-54. IEEE, 2018.
- [24] Wu, Han, Shengqi Yang, Zhangqin Huang, Jian He, and Xiaoyi Wang. "Type 2 diabetes mellitus prediction model based on data mining." *Informatics in Medicine Unlocked* 10 (2018): 100-107.