Journal of Theoretical and Applied Information Technology
15th February 2021. Vol.99. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE NEED FOR DOMAIN EXPERT IN DATA MINING APPLICATION

[1]SEUN EBIESUWA, [2]TAYO OMOLARA, [3]ADEKUNLE A. YINKA, [4]OHWO ONOME BLAISE, [5]ADESINA ADIO

[1]Doctor, Department of Computer Science, Babcock University, Ilishan-Remo, Nigeria
[2]Doctor, Department of Computer Science, Babcock University, Ilishan-Remo, Nigeria

[3]Professor, Department of Computer Science, Babcock University, Ilishan-Remo, Nigeria
[4]Student, Department of Computer Science, Babcock University, Ilishan-Remo, Nigeria
[5]Doctor, Department of Basic Sciences, Babcock University, Ilishan-Remo, Nigeria

E-mail: [1]ebiesuwao@babcock.edu.ng, [2]omolaratayo6@gmail.com, [3]adekunlea@babcock.edu.ng, [4]ohwo0247@pg.babcock.edu.ng, [5]adioa@babcock.edu.ng

ABSTRACT

In this paper we have focused a variety of Data mining techniques, approaches and applications areas which has been helpful in the important field of data mining Technologies. Organizations are operated from various geographical locations, and such operation may generate large volumes of data. The data is warehoused and used by corporate decision makers to take strategic decisions in the significant business value. Thus, improving the effectiveness of managerial decision-making. Based on the aforementioned, a comparative analysis was conducted to compare data mining techniques in various domains. To examine the need for domain expert. The value of such domain expert is easily recognized. This has drastically changed the areas of science and engineering.

Keywords: *Data Mining, Business Intelligence, Big Data, Knowledge Discovery in Database, Domain Knowledge.*

1. INTRODUCTION

In the 21st century, the increasing use of different technologies by human in their daily activities, generates large volumes of data in different fields. The data can be in various forms such as documents, graphical, video, audio, such that the proper action can be taken; analyze, manage and make a decision. And the data can be made available upon the customer's request and make the better decision. This technique is called data mining [1]. In organizations, business intelligence aims to analyze merged data in the data warehouse, a unified repository for all the data collected by the various systems, to enables business executives to organize, understand and make strategic decisions. Big data is stored in a distributed file system, allowing agile management of large volumes of data. The objective of big data is to find the knowledge from the data, which is then applied to improve any type of process. The concept of big data as a tool that facilitates the advancement of business intelligence processes; offering deep analysis and a global vision of the data. Business Intelligence provides a more structured experience [2].

1.1 History of Data Mining

The information age, has affected every sphere of human life, for example education (E-learning), medical, banking, sports, business. This resulted into large volumes of data stored in various formats. To take complete advantage of data, a tool for automatic summarization of data, extraction of the information stored, and the discovery of patterns in raw data is required. With the large volumes of stored data, it is increasingly important, to develop powerful tool for analysis, understand and extraction knowledge that could help in decision-making. This gave rise to Data mining, the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses [3]. The field of data mining has seen enormous success from the

inception, in terms of wide-ranging application achievements and in terms of scientific advancement and understanding. Data Mining though seen as a new technology is a discipline with a long history. It starts with the early Data Mining methods Bayes' Theorem (1700`s) and Regression analysis (1800`s) which were mostly identifying patterns in data.

Increasing power of technology and complexity of data sets has led Data Mining to evolve from static data delivery to dynamic and proactive information deliveries; from tapes and disks to advanced algorithms and massive databases. In the late 80`s Data Mining term began to be known and used within the research community by statisticians, data analysts, and the management information systems (MIS) communities. By the early 1990`s, data mining was recognized as a sub-process or a step within a larger process called Knowledge Discovery in Databases (KDD) [4]. The most commonly used definition of KDD is "The non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [4].

The popularity of data mining escalated notably in the 1990`s, with the help of dedicated conferences, in addition to the fast increase in technology, data storage capabilities and computers` processing speeds. It was also possible for organizations to keep data in computer readable form and processing of large volumes of data using desk top machines were not far from reality. By the end of 1990`s, data mining was already a well-known technique used by the organizations after the introduction of customer loyalty cards. This opened a big door allowing organizations to record customer purchases and data, the resulting data could be mined to identify customer purchasing patterns. The popularity of data mining has continued to grow rapidly over the last decade.

1.2 Knowledge Discovery in Database (KDD)

Data mining, also called knowledge discovery, is the process of analyzing data from different perspectives and summarizing it into useful information [5]. The following sub-processes take part in the KDD process [6]:

1. Understanding of the application and identifying the goal of the KDD process
2. Creating a target data set
3. Data cleaning and pre-processing
4. Matching the goals of the KDD process (step 1) to a particular data-mining method.
5. Research analysis and hypothesis selection
6. Data mining: Searching for patterns of interest in a particular form, including classification rules, regression, and clustering
7. Interpreting mined patterns
8. Acting on the discovered analysis

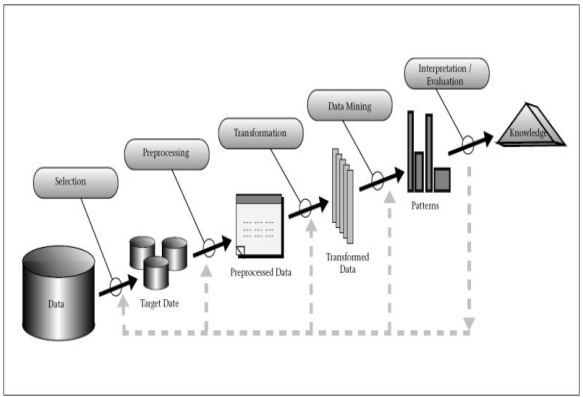

Figure 1: Knowledge discovery Process
Source: [7]

1.3 Types of Data Mining Techniques

Data mining functions include clustering, classification, prediction, and associations [8].

1.3.1 Classification rule mining algorithms

Classification deals with the construction of classifiers "that can be applied to unseen" data so as to categorized such data into groups (classes). As such classification and clustering are quite similar. The difference between them, however, is that classification requires pre-labelled training data from which the classifiers can be built. As such classification is sometimes referred to as supervised learning while clustering is considered to represent unsupervised learning. There are many methods used in data mining for classification like fuzzy logic, Artificial Neural Networking, Decision trees. Classification is one of the most frequently studied problems by Data mining and machine learning (ML) researchers. It consists of predicting the value of a (categorical) attribute (the class) based on the values of other attributes (the predicting attributes) [9].

1.3.1.1 Overview of classification rule mining algorithms

1. C4.5 Algorithm

Systems that construct classifiers are one of the commonly used tools in data mining. Such systems collect as input a collection of cases, each belonging to one of a small number of classes and are described by its values for a fixed set of attributes and output a classifier that can accurately predict the class to which a new case belongs [10].

2. Support Vector Machines (SVM)

SVM was first introduced by Vapnik and has been a very effective method for regression, classification and general pattern recognition. It is considered a good classifier because of its high generalization performance without the need to add a priori knowledge, even when the dimension of the input space is very high [10].

3. K Nearest Neighbors

The k-nearest neighbors' algorithm is amongst the simplest of all machine learning algorithms. [10] classified an object by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors. *k* is a positive integer, typically small. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

4. Naïve Bayes

This is another classification method; it is important for several reasons namely [10];

1. It is very easy to construct, not needing any complicated iterative parameter estimation schemes. This means it may be readily applied to huge data sets.
2. It is easy to interpret, so users unskilled in classifier technology can understand why it is making the classification it makes.

1.3.1.2 Related works on classification rule mining algorithms

[11] did a comparative analysis of classification algorithms on different datasets using Weka, the algorithms used were J48 and multilayer perceptron algorithms and it was discovered that multilayer perceptron was a better algorithm. [12] proposed a new rule-based algorithm for classifying and predicting uncertain datasets. They proposed new approaches for deriving optimal rules out of highly uncertain data, pruning and optimizing rules, and class prediction for uncertain data. The algorithm proposed followed the new paradigm of directly mining uncertain datasets.

1.3.2 Clustering rule mining algorithm

Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups. Representing data by fewer clusters necessarily loses certain fine details (akin to lossy data compression) but achieves simplification. It represents many data objects by few clusters, and hence, it models data by its clusters. Data modelling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a data concept. Therefore, clustering is unsupervised learning of a hidden data [13].

1.3.2.1 Overview of clustering rule mining algorithms

1. Chameleon Clustering Algorithm

This is a hierarchical clustering algorithm that uses dynamic modeling to determine the similarity between pairs of clusters. In chameleon, cluster similarity is assessed based on how well-connected objects are within a cluster and the proximity of clusters. In this algorithm two clusters are merged if their interconnectivity is high and they are close [13].

2. Balanced iterative reducing and clustering using hierarchies (BIRCH)

This is designed for clustering a large amount of data bases. It is recognized as the first clustering algorithm proposed in the database area to handle noise effectively. BIRCH incrementally constructs a CF (Clustering feature) Tree. It is a multiphase clustering technique in which phase 1 scans database to build an initial in- memory CF tree and phase 2 uses an arbitrary clustering algorithm to cluster the leaf nodes of the CF tree [13].

3. Clustering algorithm based on randomized search (CLARANS)

This is a clustering process which can be presented as searching a graph where every node is a potential solution. It draws sample of neighbors dynamically. It is more efficient and scalable. The main aim is to identify spatial structure that may be present the data. It can handle not only point objects, but also polygon objects efficiently [13].

1.3.2.2 Related works on clustering rule mining algorithms

Clustering has always been used in statistics and science [14] presented a new efficient which combines both the cluster concept and decomposition of larger candidate item sets, while proceeds from mining the maximal large item sets down to large 1-item-sets, named cluster-decomposition association rule (CDAR). [5] also developed a hybrid approach using rule induction and clustering techniques in terms of accuracy and processing time in Data Mining they used induction algorithms and clustering as a hybrid approach to maximize the accurate result in fast processing time. [15] presented a novel clustering algorithm for log file data sets which helps one to detect frequent

patterns from log files, to build log file profiles, and to identify anomalous log file lines.

1.3.3 Association rule mining algorithm

Association rule aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control [16]. Over the last several years, the problem of efficiently generating large numbers of association rules has been an active research topic in the data mining community. Mining for association rules can help in business decision making, and the development of customized marketing programs and strategies [17]. Association is the discovery of association relationships or correlations among a set of items. This problem was introduced in [18]. The selection of association rule is based on support and confidence. The confidence factor indicates the strength of the implication rules; while the support factor indicates the frequencies of the occurring patterns in the rule [16]. An example of an association rule is: "35% of transactions that contain bread also contain milk; 5% of all transactions contain both bread and milk". In this case, the confidence of the rule is 35%, and 5% the support of the rule [19].

1.3.3.1 Overview of association rule mining algorithms

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually sub divided into two sub problems. One is to find those item-sets whose occurrences exceed a predefined threshold in the database; those item-sets are called frequent or large item-sets. The second problem is to generate association rules from those frequent item-sets with the constraints of minimal confidence. Suppose one of the large item-sets is *Lk, Lk = {I1, I2, …, Ik},* association rules with this item-sets are generated in the following way: the first rule is *{I1, I2, … , Ik-1}! {Ik},* by checking the confidence of this rule the level of interestingness can be determined. Then other rules are generated by deleting the last items in the antecedent and inserting it to the consequent, the new rules are checked to determine the interestingness of them. Those processes iterated until the antecedent becomes empty. Since the second sub problem is quite straight forward, most of the researches focus on the first sub problem. The first sub-problem can be further divided into two sub-problems: candidate large item-sets generation process and frequent item-sets generation process. We call those item-sets whose support exceed the support threshold as large or frequent item-sets, those item-sets that are expected or have the hope to be large or frequent are called candidate item-sets [16]. In many cases, the algorithms generate an extremely large number of association rules, often in thousands or even millions. Further, the association rules are sometimes very large. It is nearly impossible for the end users to comprehend or validate such large number of complex association rules, thereby limiting the usefulness of the data mining results. Several strategies have been proposed to reduce the number of association rules, such as generating only "interesting" rules, generating only "non-redundant" rules, or generating only those rules satisfying certain other criteria such as coverage, leverage, lift or strength.

1.3.3.2 Review of related works on association rule algorithms

[20] proposed an association rule algorithm that uses all existing information between database passes to avoid checking the coverage of redundant sets. The algorithm gives clear empirical improvement when compared against the previous results, and it is simple to implement. [21] showed that based on experimental result FP Growth algorithm is better than traditional Apriori algorithm in case of number of rules, CPU time and minimum support. [22] discussed six association rule mining algorithms with their example: AIS, SETM, Apriori, Apriori tid, Apriori hybrid, FP-growth. Comparison was done based on the above performance criteria. Each algorithm has some advantages and disadvantages. From the above comparison they concluded that, FP-growth performs better than all other algorithms. [23] dealt with the algorithmic aspects of association rule mining. They compared the most important algorithms and analyzed their performance based on runtime experiments and theoretic considerations.

1.4 The Data Mining Task

There are different types of data mining tasks, which is classified based on the use of the data mining result [1]:

1. **Exploratory Data Analysis:** This technique is interactive and visual to the user. This allows the user to search the data repositories. Then data mining analyzes the data.
2. **Descriptive Modeling:** this technique describes all the available data in the repositories. it includes models for overall probability distribution of the data, partitioning of the p-dimensional space into groups and models

describing the relationships between the variables.

3. **Predictive Modeling:** This technique allows the value of one variable to be predicted from the known values of another variables.
4. **Discovering Patterns and Rules:** This task is primarily used to find the hidden pattern as well as to discover the pattern in the cluster. The aim of this task is "how best we will detect the patterns".
5. **Retrieval by Content:** This technique finds pattern similar to the pattern of interest in the data set.

1.5 Data Mining Life Cycle

The life cycle of a data mining project consists of six phases. The sequence of the phases is not rigid, as movement between different phases is always required and depends on the outcome of each phase [3] [1]. The main phases are:

1. **Business Understanding:** This focuses on understanding the objectives and requirements of the project from a business perspective. Then, it is converting into a data mining problem definition and an initial plan designed to achieve the objectives.
2. **Data Understanding:** It starts with an initial data collection, to get familiar, identify data quality problems, discover first insights or detect interesting subsets to form hypotheses for hidden information.
3. **Data Preparation:** All the different data sets are collected and a variety of the activities based on the initial raw data are constructed.
4. **Modeling:** various modeling techniques are selected and applied and their parameters are calibrated to optimal values.
5. **Evaluation:** Here the model is thoroughly evaluated and reviewed. The process of construct the model is executed, to be certain it works properly and can achieve the business objectives. Then, a decision on the use of the data mining results should be reached.
6. **Deployment:** The purpose of the model is to increase knowledge of the data. The knowledge gained will need to be organized and presented in a way that the customer can use it. This can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise.

2 LITERATURE REVIEW

This section highlights some of the application of data mining in various aspects of human endeavour. The computer science field of data mining, with the purpose of extracting meaning from data, this, expressly looking for patterns in historic observations and predicting future behavior. [24] considered the techniques, and interpretation of data mining, with specific focus on its application in audiology. Modern hearing instruments contain data-logging technology to record data separate from the audio stream and how the signal processing was consequently operating; combined with patient details. The variety of data generated lends itself to a data mining approach. [25] investigated how to objectively evaluate the financial risks existing in enterprise management and how to timely warn them. The results show that the uncertainty of the technology is very high, and the excellent performance of data mining technology in the study of uncertainty theory links the two closely. With the rapid increase of data storage capacity, massive data processing and massive data calculation has become an important problem in the field of data mining. Cloud computing is good at dealing with large-scale data and large-scale computing. If the data mining algorithm can be cooperated to the cloud computing platform, the large computational problems in the field of data mining will be solved [26]. For improving the speed of character recognition, [27] applied data mining technology to character recognition under the condition of big data architecture. [28] studied the specific application of data mining technology in economic analysis of folk art from the two aspects of k-means clustering algorithm and weighted CADD algorithm respectively. [29] investigated the risk factors that could predict intraoperative hemodynamic instability (IHD) during pheochromocytoma surgery by data mining. [30] used 705,747 POI (Point of Interest) to conduct simulation analysis of western cities in China by mining the data of online maps. Through kernel density analysis and spatial correlation index, the distribution and aggregation characteristics of different types of POI data in urban space were analyzed and the spatial analysis and correlation characteristics among different functional centers of the city were obtained. [31] summarized the problems of the data information existing in the application of ERP system in coal enterprises, and point out the importance of the application of data mining technology. Through prediction and comparison on turnover of nearly five years of a part of the coal enterprises using the data mining technology, it was found that the application of data mining technology in the ERP enterprises is of great significance. In order to better study the portfolio optimization problem in this era of big data, [32] proposed a method of integrating data mining

technology with portfolio optimization problem, and makes full use of data mining technology. In a large number of portfolio optimization problem data, the relevant data are accurately analyzed and efficiently managed, so as to reasonably optimize the portfolio Research. It is found that the method proposed in this paper is of great significance to the application of data mining technology in portfolio optimization. Information means such as database and education cloud are effectively used by [33] to make an in-depth exploration of the music curriculum model based on cloud computing and data mining technology to provide higher quality information services to schools, teachers and students.

3 METHODOLOGY

In this section, a comparative analysis of data mining techniques across different domain is presented. This helps to provide a basic understanding of the extent to which data mining is being applied. And also, to establish a view that domain expert is needed when considering the data mining technique to us

Table 1: A Comparative Analysis of Data Mining Techniques across different Domains

Author(s)/ Year	Title	Domain	Methodology	Result(s)	Gap(s)
[34]	Application of Data Mining to a Large Hearing-Aid Manufacturer's Dataset to Identify Possible Benefits for Clinicians, Manufacturers, and Users	Audiology	The process, techniques, and interpretation of data mining, with specific focus on its application in audiology.	This should be considered as a ''proof-of-concept'' of the potential for data mining of such large datasets.	The redaction, as well as our arms-length relationship with the anonymous manufacturer, has meant that the findings may be explicable by factors to which we were not privy and may be much more prosaic. Because of the extensive searching, some of the findings may exist purely by chance.
[25]	Application of Data Mining Technology in Financial Risk Analysis	Finance	Using the method of association rules interactive mining, the financial risk index of many aspects is more widely chosen.	Compared with the traditional statistical analysis method, this method is more objective and practical.	
[26]	The development and application of data mining based on cloud computing	Cloud Computing	Introduces the Basic characteristics and process of cloud computing and data mining, including	Cloud computing with its distributed computing platform provides a powerful	

			neural network method, genetic algorithm, decision tree method, statistical analysis, rough set method	computing. Thus, the combination of cloud computing and data mining has a huge advantage and potential.	
[27]	Research on application of data mining in fast character recognition based on big data	Pattern Recognition	Application of data mining technology, such as Artificial Neural Network, to character recognition under the condition of big data architecture	The proposed data mining algorithm can further improve the learning efficiency and improve the performance of the time	The main difficulties in the research of license plate recognition are the low quality of vehicle license plate image, the small proportion of license plate area, and the distortion of characters to be recognized.
[28]	Application of data mining technology in regional economic analysis	Economics	The specific application of data mining technology in economic analysis of folk art from the two aspects of k-means clustering algorithm and weighted CADD algorithm respectively	When using weighted CADD algorithm, it can be found that it divides the regions rich in mineral resources in China into the same category, while k-means clustering algorithm is more general, but the clustering effect is also correct and has scientific basis.	According to the actual data, it can be concluded that the clustering accuracy will decrease as the dimension increases. Only after dimension reduction can the clustering result be called accurate.
[29]	Application of data mining for predicting hemodynamic instability during	Medicine	Relief-F was used to select the most important features. The accuracies of	Random forest had the highest AUC and accuracy values of 0.8636 and	Relatively small sample size and number of events included in this study, may lead to

	pheochromocytoma surgery		seven data mining models (CART, C4.5, C5.0, and C5.0 boosted), random forest algorithm, Naive Bayes and logistic regression were compared. The cross-validation, hold-out, and bootstrap methods were used in the validation phase	0.8509, respectively. Then, we improved the random forest algorithm according to the classification of imbalanced data. Improved random forest model had the highest specificity and precision among all algorithms, including relatively higher sensitivity (recall) and the highest f1-score integrating recall and precision.	underestimation of its predictive effect.
[30]	Research on the Simulation Application of Data Mining in Urban Spatial Structure	Urban Planning	Through kernel density analysis and spatial correlation index, the distribution and aggregation characteristics of different types of POI data in urban space were analyzed and the spatial analysis and correlation characteristics among different functional centers of the	Data mining of Internet of things (IOT) has good adaptability in city simulation and will play an important role in urban research in the future	The multi-center development of the city is still in its infancy, the differences between the centers are obvious, and there is a big gap from the multi-centers of the city.

			city were obtained.		
[31]	Application of Data Mining in the Coal Enterprises in Enterprise Resource Plan	Enterprise Resource Planning	Apply data mining technology to establish the system of enterprise ERP application system.	Through prediction and comparison on turnover of nearly five years of a part of the coal enterprises in China by using the data mining technology, the application of data mining technology in the coal enterprise ERP is feasible and very fruitful.	
[32]	Application of Data Mining Technology in Portfolio Optimization	Risk Management	In order to meet the new requirements of the new era for portfolio optimization, this paper puts forward the combination of data mining technology and investment	The method proposed is of great significance to the application of data mining technology in portfolio optimization.	
[33]	Application of Data Mining Technology in Music Curriculum Resources	Music	The music curriculum model based on data mining technology	It has successfully solved the problems of uneven distribution of music education resources, low sharing degree, simple and boring contents, and intelligent adjustment unavailable, thereby providing	

				higher quality information services to schools, teachers and students.	

4 DISCUSSION OF FINDINGS

From the literature reviewed on the various application areas of data mining use various forms of data ranging from text to images and are stored in databases and/or repositories. And various data mining techniques are used to extract the patterns and knowledge. Data selection and data mining technique are important task and knowledge of the domain is needed. With several attempts to design and develop a general data mining system failed. Hence, a domain expert's assistant is mandatory in every domain. Such that, effectively applying the knowledge of the domain experts and use of data mining systems can generate the required knowledge.

The domain experts are required to carry out the following functions based on a specific problem domain [1]:

- Determine the various data formats that should be collected,
- selection of specific data for data mining,
- cleaning and transformation of data,
- extracting patterns for knowledge generation and
- interpretation of the patterns and knowledge generation.

From the literature reviewed, most of the domain specific data mining applications show accuracy above 90%. The intelligent interfaces and agents, to a considerable level, make the application general but not without limitations. The domain experts play a vital role by aiding in the decision making at different stages in data mining. The domain specific applications are aimed at extracting specific knowledge by considering user requirements and other context parameters guiding the system. The results from the domain specific applications are more accurate and useful. Therefore, it is concluded that the domain specific applications are more specific for data mining. Consequently, the design and develop a data mining system, that can dynamically work in any domain would be daunting task.

5 CONCLUSION

Data mining has standing regarding finding the patterns, forecasting and discovery of knowledge, in various domains. Data mining techniques and algorithms are the tools to finding the patterns to decide upon the future trends in businesses to grow. Data mining has wide application domain almost in every industry where data is generated. Consequently, data mining is considered one of the most important borders in database and information systems and one of the most promising interdisciplinary developments in Information Technology.

REFERENCES:

[1] N. Padhy, P. Mishra and R. Panigrahi, "The Survey of Data Mining Applications And Feature Scope," *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT),* vol. 2, no. 3, pp. 43-58, 2012.

[2] T. Velásquez-Pérez, J. A. Camargo-Pérez and E. L. Quintero-Quintero, "Application of data mining as a tool in computer science," in *6th International Week of Science, Technology and Innovation (6th IWSTI)*, 2020.

[3] D. .. Larose, Discovering Knowledge in Data Mining: An Introduction to Data Mining, New Jersey: John Wiley & Sons, Inc, 2005.

[4] U. Fayyad and P.-S. Gregory, "From Data Mining to Knowledge Discovery in Databases," *American Association for Artificial Intelligence,* pp. 37- 54, 2006.

[5] s. Kapil and V. Sheveta, "A hybrid Approach Using Rule Induction And Clustering Techniques In Terms Of Accuracy And Processing Time In Data Mining," *International Journal of Engineering and Advanced Technology (IJEAT),* 2013.

[6] M. Venkatadri and R. Lokanatha, "A Review on Data mining from Past to the Future,"

International Journal of Computer Applications, p. 0975 – 8887, 2011.

[7] B. M. Ramageri, "DATA MINING TECHNIQUES AND APPLICATIONS," *Indian Journal of Computer Science and Engineering,* vol. 1, no. 4, pp. 301-305, 2011.

[8] U. Irshad, "Data Mining Algorithms And Medical Sciences," *International Journal of Computer Science & Information Technology (IJCSIT),* pp. 56-69, 2010.

[9] T. Ravindra and I. Avinash, "Credit Card Fraud Detection Using Hidden Markov Model and Its Performance," *International Journal of Advanced Research in Computer Science and Software Engineering,* 2016.

[10] R. Kumar and R. Vemar, "Classification Algorithms for Dat Mining," *International Journal for Innovations in Engineering and Technology,* pp. 7-14, 2012.

[11] A. Rohit and Suman, "Comparative Analysis of Classification Algorithms on Different Datasets using WEKA," *International Journal of Computer Applications,* 2012.

[12] Q. Biao, X. Yuni, P. Sunil and T. Yicheng, "A Rule-Based Classification Algorithm for Uncertain Data," *International Conference on Data Engineering,* pp. 1633-1640, 2013.

[13] G. Pooja and M. Jena, "Comparative Study of Different Clustering Algorithms for Association Rule Mining," *International Journal of Computer Science and Engineering Technology,* pp. 592-595, 2013.

[14] M. Luc and L. Kaufman, The Interpretation of Analytical Chemical Data using Cluster Analysis, Wiley and Sons Inc, 2003.

[15] R. Vaarandi, "Data Clustering Algorithm for Mining Patterns From Event Logs," *IEEE Workshop on IP Operations and Management,* pp. 7803-8199, 2003.

[16] K. Sotiris and K. Dimitris, "Data Preprocessing for Supervised Learning," *International Journal for Computer Science,* pp. 111-126, 2006.

[17] P. S. Patil, J. V. Aghav and V. Sareen, "An Overview of Classification Algorithms and Ensemble Methods in Personal Credit Scoring," *International Journal of Computer Science And Technology,* pp. 183-188, 2016.

[18] A. Rakesh and S. Arun, "Mining Association Rules Betwwen Sets of Items in Large Databases," in *International Conference of Management of Data*, 2003.

[19] A. Akshai, "Electricity price forecasting in deregulated markets: a review and evaluation.," *International Journal Electricity Power Energy System,* pp. 13-22, 2009.

[20] M. Heikki and H. Toivonen, "Efficient Algorithms for Discovering Association Rules," *AAAI Technical Report,* 2005.

[21] P. Banu and U. Vekheera, "Enrinching the Efficiency of Association Rule Mining using Enhanced IFP-Ggrowth Algorithm," *International Journal of Advanced Research in Computer Science and Software Engineering,* 2013.

[22] K. A. Trupti and C. V. Santosh, "An Overview of Association Rule Mining Algorithms," *International Journal of Computer Science and Information Technologies,* pp. 927-930, 2013.

[23] H. Jochen and G. Ulrich, "Algorithms for Association Rule Mining- a general survey and comparison," *Inernational Journal for Computer Science,* pp. 145-160, 2000.

[24] J. C. Mellor, M. A. Stone and J. Keane, "Application of Data Mining to ‘‘Big Data’’ Acquired in Audiology: Principles and Potential," *Trends in Hearing,* vol. 22, pp. 1-10, 2018.

[25] M. Jin, Y. Wang and Y. Zeng, "Application of Data Mining Technology in Financial Risk Analysis," *Wireless Pers Communication,* pp. 1-15, 2018.

[26] Z. Jun, "The development and application of data mining based on cloud computing," in *First International Conference on Advanced Algorithms and Control Engineering*, 2018.

[27] Y. Zheng, W. Song and L. Shen, "Research on application of data mining in fast character recognition based on big data," *Concurrency Computation Practical Experiment,* pp. 1-12, 2018.

[28] S. Zhu, "Application of data mining technology in regional economic analysis," in *Journal of Physics: Conference Series*, 2019.

[29] Y. Zhao, L. Fang, L. Cui and S. Bai, "Application of data mining for predicting hemodynamics instability during pheochromocytoma surgery," *BMC Medical Informatics and Decision Making,* pp. 1-13, 2020.

[30] J. Zhang, X. Sui and X. He, "Research on the Simulation Application of Data Mining in Urban Spatial Structure," *Journal of Advanced Transportation,* pp. 1-9, 2020.

[31] R. Zhong and H. Wang, "Application of Data Mining in the Coal Enterprises in ERP," in *Journal of Physics: Conference Series ICNISC 2020*, 2020.

[32] G. Hu and H. Liu, "Application of Data Mining Technology in Portfolio Optimization," in *Journal of Physics: Conference Series BTCS 2020*, 2020.

[33] L. Zou, "Application of Data Mining Technology in Music Curriculum Resources," in *Journal of Physics: Conference Series ICAIIT 2020*, 2020.

[34] J. C. Mellor, M. A. Stone and J. Keane, "Application of Data Mining to a Large Hearing-Aid Manufacturer's Dataset to Identify Possible Benefits for Clinicians, Manufacturers, and Users," *Trends in Hearing,* vol. 22, pp. 1-15, 2018.

[35] C. Ming-Syan and H. Jiawei, "An Overview from a Database Perspective Database Perspective," *IEEE Transactions on Knowledge and Data Engineering,* pp. 866-883, 2006.

[36] S. Samaneh, Z. Zahra, E. A. Reza and H. M. Amir, "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective," pp. 1-26, 2016.