

# Challenges and solutions for manufacturing products using advanced enterprise data mining approaches

M.Srimathi<sup>1</sup>, Dr. R.Umarani<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Dravidian University, Kuppam, (A.P), India

<sup>2</sup>Associate Professor, Department of Computer Science, Sri Sarada College for Women, Salem, (T.N), India

<sup>1</sup>srimathiphdresearch5@gmail.com, <sup>2</sup>umaranihd123@gmail.com

## Abstract

**Objectives:** The major intention of this investigation is to increase the accuracy of manufacturing enterprises using efficient data mining techniques. This scenario also intends to accomplish the higher performances by reducing the challenges and producing the potential solutions.

**Methods:** In this research, k-nearest neighbor (kNN), Naïve bayes (NB), C4.5 and J48 algorithms are used to evaluate the medical database, transactional database and agricultural database. This scenario considers the feature extraction and feature selection to improve the accuracy of system.

**Findings:** The algorithms are used in this work namely kNN, NB, C4.5 and J48 algorithms are used to improve the manufacturing performances. The experimental results are shown that the J48 classification algorithm is superior to other algorithms such as kNN, NB and C4.5. These techniques are improved in terms of all performance metrics called accuracy, precision and recall values.

**Application/Improvements:** The findings demonstrate that accuracy performance is improved significantly by using J48 algorithm in this scenario. The data mining algorithms are used to reduce the challenges tremendously and provide optimal solutions for manufacturing enterprises.

**Keyword:-** Manufacturing, Enterprise data, Data mining algorithms, Information discovery

## 1. Introduction

Data mining is the process of discovery of hidden patterns from the huge amount of data. The diverse techniques and algorithms of data mining are classification, clustering, association rule mining, sequence analysis, time sequence and outlier analysis. The efficient information mining method should be selected according to the particular study area. The sequence of data mining can be evocative knowledge or analytical knowledge. Enterprise systems reflect the typical operations of huge corporations. The enterprise is described as a large industry relationship which survives either to produce some manufactured goods, or to present some types of service as part of their profit searching behavior. The manufactured goods could be agricultural, material, house ware substances, transport associated, games possessions, and any other engineered manufactured article. The service produced could be healthcare, economics, effectiveness, telecommunication, transportation, preservation and clean. The enterprise in the trade of manufacture some artifacts is an industrialized enterprise and the enterprise in the production of producing service is a service enterprise.

Data mining which extracts the products or items and discovering predictive information in single huge dimensional data ware house. Enterprise started via using processors to supervision data linking to daily trade business, thus, saving expensive manpower and improving the information precision. Data mining and knowledge discovery in large databases is focused a prominent amount of investigation. There is an emergency requirement for a new generation of computational theories and tools for supporting human in mining important knowledge from the speedily increasing amount of digital data. These advanced theories and tools are theme of the promising field of knowledge discovery in databases. A huge amount of the present interest in KDD is the outcome result of the social media concerned surrounding successful KDD applications. In marketing, the major application is database advertising systems, which examine client database to identify the diversity client groups and predict their activities. In finance, several companies utilize data mining for investment and the methods of data mining is used for providing better decision making.

Similarly in telecommunications it is useful for manufacturing the telecommunication equipments and networks. KDD is having several applications in healthcare such as analysis, patient dealing, administration of chronic sickness, forecast of patients at threat for particular illness, and in public health. Extracting data from raw information is an enormously essential and heavy procedure in today's data driven world. Enterprises nowadays rely on a set of computerized tackle for information detection to increase commerce imminent and aptitude.

Numerous branches of knowledge discovery tools has been expanded to assist today's aggressive big business advertise succeed in the age of information. World's electronic financial system has also improved the difficulty on enterprises to adjust to such novel production background. Major tools for receiving knowledge data from these huge quantities are programmed extracting tools, particularly communication data mining, text mining, and web mining.

Business applications are based on the extraction of the software solutions. Extracting tools are nowadays an important element of enterprise decision-making and risk organization. Attaining information data via extraction is referred to as Business Intelligence BI. Enterprise datasets are raising speedily, recognition to utilize of Knowledge Systems IS, and data warehousing. The idea of e-agricultural is explained as the incorporation and employment of information technology in farming associated procedures. Farming is a conventional trade and operations concerned in manufacture have fundamentally stayed behind consistent in natural history. Conventional approaches of farming integrated modest to no employ of information technology, in spite of its possible for considerably rising procedures good organization. The integration of information technology into agricultural occupies the incorporation of various expertise, along with each accomplished of optimistically impacting the effectiveness of agricultural behavior, thus supporting sustainable agricultural practice.

In [1] G Geethakumari and agrima srivastava discussed enterprise data security using data mining algorithms. The enterprise data which can be retail, finance, health and manufacturing information is to be analyzed. The emerging security has been identified using unsupervised machine learning algorithm and Bayesian network algorithm. In [2] Muhammad Shahbaz and Syed Athar Masood recommended data mining methodology to provide the solutions for manufacturing challenges. This scenario is used to individual enterprise and to an extended enterprise to ensure profits of data mining method and to share the discovered knowledge among enterprises. However it has issue with noisy training data hence it produces inconsistent results.

Antima arora and Nidhi mittal are suggested data mining methods for online analytical processing. The data mining algorithms are making easy the managers through providing them along with related knowledge [3]. John leveille discussed the issues of OLAP for medical healthcare information. This scenario is recommended the quality of health care information for patients on time. In [4] neural network algorithms are introduced for recognizing the unclassified genes and sequences. It is also used to provide the new developments in data mining applications. Sami ben jabeur discussed classification of business failure by using data mining methods. This scenario utilized the approaches named as partial least square discriminant analysis and support vector machine. The prediction of enterprises through the support vector machine is more accuracy. But however it does not able to execute the huge number of dataset with huge failures [5].

In another research [6], the support vector machine algorithm is applied to evaluate the data classification. The numbers of datasets are taken for the computation and SVM algorithm provides higher classification results. This algorithm is able to handle high dimensional dataset but this scenario suffers from time complexity. Oleg chertov is recommended [7] enterprise architecture model which is used for analyzing the transactional database. This scenario is constructed enterprise architecture to enable searching for statistical information. In [8] Wei-sen chen and yin kuan du suggested the data mining algorithms such as neural network. This scenario is used to classify the financial distress using neural network model efficiently. In this method, the accuracy error rate is reduced significantly and achieve better prediction accuracy rate. This research is introduced artificial intelligence (AI) method which is more applicable technique to classify the powerful financial distress of a company.

In [9], the method developed to design retail promotions, informed by product associations observed in the same groups of customers. It used the Clustering and association rule find to identify customer behavior. It can easily predict the sales. The customer with similar purchasing behavior are first grouped by means of clustering techniques such as K-means method and for each cluster an association rule (Apriori algorithm) to identify the products that are brought together by the customers. Prabha Dhandayudam et al [10] attempted to improve clustering algorithm for segmenting the customer using RFM (Recency, Frequency, and Monetary) values. Then the performance of the algorithm compared with other traditional techniques such as K-means, single link and complete link. RFM is very effective method for customer segmentation. For segmenting the customers, the attribute R, F and M are used as three in clustering techniques. For finding the distance between from each objects to all other object, here Manhattan distance used and store it in distance matrix. It experimented with real data set of the customer transaction details are used for clustering.

In [11], Clifton phua and Vincent lee discussed data mining algorithms applications in the field of discovery of fraud research. This scenario used fuzzy neural network algorithm to detect effectively the frauds in the system. It is also used to speed up the process and cost of the scenario is in expensive. However it has issue with label data information and hence performance of the scenario is reduced. In [12], Chris rygielski suggested data mining

methods for dealing the consumer relationship organization. This algorithm is improving the knowledge and understanding of customers, items or services and transactional information to progress the decision making. However it has issue with privacy in this scenario. [13] Faced along with the increasing issues caused through worldwide contestants, Europe's manufacturing segment has wanted to new by using computerization resolutions that offer producers with the means to enhance their efficiency via increasing the superiority and reliability of their foodstuffs when decreasing their operating expensive. Additional to this, producers also utilize computerization resolutions in order to present fitness and security profits, as they eliminate the necessitate for persons to work in dangerous or insecure surroundings, as well as environmental profits, in the form of more power resourceful manufacture processes. In the case of robotics, which entails the plan, creation and accomplishment of machinery to execute responsibilities conventionally performed through humans, developed processes profits most from superior flexibility. Robots are flexible in that they could simply modify their purpose in order to assemble the difficulty of the producer or the consumer. In doing so, robots fundamentally execute three responsibilities, as they "sense" via illustrating on ecological stimulus, then "think" via using predetermined algorithms for preparation and finally "act" using the robots' end-effectors to select also place an article or fuse two substances jointly. However, such progresses in dimension methodologies and robotics, the implementation of computerization resolutions are controlled by two factors. It is such as their specific reliance on the automotive segment in Europe and their inadequate relevance by small and medium size enterprises, who consider it unsuitable to incorporate capital-intensive robots into their small-scale industrialized processes.

## 2. Materials and Methods

### 2.1. Enterprise data mining for transactional database

The transactional database, agriculture database and medical database are considered in this scenario to evaluate the manufacturing challenges and solutions by suing effective data mining algorithms. In transactional database, the transaction data refers to data obtained when a transaction takes place, such as product name, quantity, location and time of purchase. Transaction database is a collection of documentations which demonstrating the transactions each with timestamp, identifier and item sets. By using an effective data mining techniques, can provide proper decision making for business transactions. Communicating with consumers consistently means business should store the transaction records and responses in an online system which is accessible to employees who know how to communicate with it. The significance of founding close consumer relationships is identified [14]. In [15] suggested that consumer association management is suitable only for directing associations among business and consumers. A faster investigation discloses that the transactional database is more crucial for business clients. In business-to-business (B2B) situation, an incredible quantity of knowledge is substituted on a standard foundation. For instance, business is more frequent, custom deals are more different, and pricing schemes are more difficult. Consumer relationship assists soft the process while numerous representatives of vendor and purchaser interact and collaborate.

Personalized catalogues, adapted production portals, and targeted manufactured offers can shorten the procurement procedure and progress efficiencies for companies. Mail alerts and novel invention information tailored to various roles in the consumer business could assist the efficiency of auctions field. Reliance and influence are improved if objective academic statements news is distributed to the appropriate individuals. All of above mentioned are assumed the advantages of customer relationship management. For the transactional database, the algorithms suggested are such as k-nearest neighbor, naïve bayes, c4.5 and J48 algorithms. These algorithms are used to estimate the performances of manufacturing in transactional database. And also it is used to evaluate the highest performance metrics of each algorithm more effectively.

### 2.2. Enterprise data mining for medical database

For medical database, M. Durairaj and V. Ranjani paying attention to evaluate a diversity of methods, algorithms and dissimilar tools and its impact on the healthcare segment. The objective of data mining appliance is becoming that information is facts, records, or manuscript which can be accessed through a computer into knowledge. The major significance of information extraction functions in medical systems is to progress a computerized tool for recognizing and broadcasting appropriate healthcare knowledge. This scenario intends to construct a comprehensive revision statement of diverse kinds of information extraction applications in the healthcare area and to condense the difficulty of the learning of the healthcare information dealings. Also provides a relative revision of special data mining functions, methods and several technologies used for mining information from database produced in the healthcare manufacturing [16].

Data mining algorithms used in healthcare manufacturing play an important task in forecast and analysis of the infections. There are a huge degree of information extraction functions are discovered in the therapeutic associated applications such as health check appliance manufacturing, prescription production and hospice organization. To determine the helpful and concealed information from the catalogue is the intention behind the appliance of information extraction. Universally data mining named information detection from the data. The information finding is a communicative process, containing through increasing a perceptive of the appliance area, choosing and generating a data set, pre-processing, information conversion. Information extraction is applied in a diversity of purposes such as advertising, purchaser association organization, manufacturing, and medication investigation, skilled forecast, web mining and mobile computing. For medical database, the algorithms are recommended such as naïve bayes, c4.5, J48 and k-nearest neighbor algorithm which evaluate the performances. These algorithms are used to provide useful information for health care organization and also provide solutions while manufacturing the medical enterprise data.

### 2.3. Enterprise data mining for agriculture database

In [17], D.Ramesh suggested that the efficient data mining techniques are used in the area of agricultural in terms of quality production with highest yield. The problem in yield prediction is resolved by using effective information extraction approaches. The methods are focused on the determination of suitable model which accomplishes the greater accuracy as well as higher generality by means of manufacture prediction capabilities. The algorithms are detecting the necessary information about farming data. The main conditions are such as specifying the raw and derived variables have a prominent impact on agricultural dataset. This is called as feature selection which chooses only important features. Training of probabilistic models is which could forecast the excellence gains as a function of recognized analytical features. This is called as classification or prediction which classifies accurately based on the specified model. Feature extraction is used to describe the association among predictive and target features. Then we have to calculate the predictive accuracy of induced models on prospect terms. The algorithms are applied in the agricultural database such as k-nearest neighbor, naïve bayes, c4.5 and J48 algorithms.

## 3. Results and Discussion

In this section the performance metrics are compared by using efficient data mining algorithms for transactional, medical and agricultural databases. The performance metrics are considered as accuracy, precision and recall values. The algorithms such as k-nearest neighbor, naïve bayes, C4.5 and J48 are used to estimate the above mentioned performance metrics. These algorithms are used to improve the performances of transactional, medical and agricultural databases. From the experimental result, we can conclude that J48 algorithm is used to improve the accuracy, precision and recall values rather than other algorithms. Hence J48 classification algorithm is superior in predicting the decision in manufacturing enterprises.

### 3.1. Accuracy

The accuracy of the classification rate is measured with the values of the True Negative, True Positive, False Positive, False negative actual class and predicted class results it is defined as follows,

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

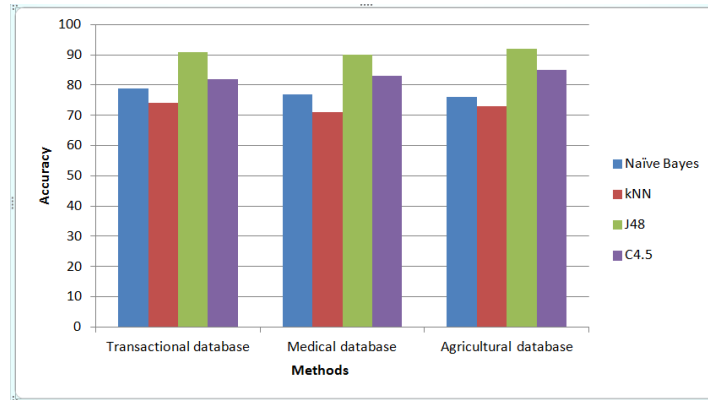
From the Figure 1 it can be proved that the J48 classification algorithm provides better result than the other algorithms by increasing the accuracy values. In this graph, methods are plotted in the x axis and the accuracy values are plotted in the y axis. The accuracy value is low by using kNN, NB and C4.5 algorithms for all enterprises. The accuracy value is significantly increased by using J48 algorithm for all enterprise databases. Thus the J48 algorithm is used to improve the performances prominently. From the experimental result, we can conclude that the J48 algorithm superior to other algorithms.

The values given in Figure 1 are illustrated as table format. The table 1 shows the different accuracy values for various algorithms and various enterprises. From the table values the J48 algorithm yields higher accuracy values compare than other algorithms. Hence the J48 is used to improve the solutions more accurately.

Table 1. Accuracy values

Enterprises	NB	kNN	J48	C4.5
Transactional database	79	74	91	82
Medical database	77	71	90	83
Agricultural database	76	73	92	85

Figure 1. Accuracy



**3.2. Precision**

Precision value is calculated based on the retrieval of information at true positive prediction, false positive. In healthcare data precision is calculated the percentage of positive results returned that are relevant.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

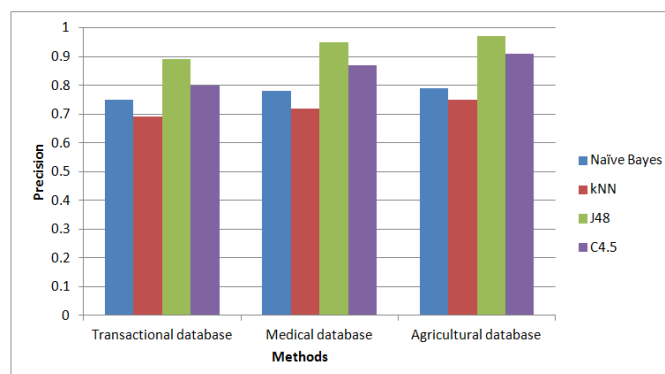
From the Figure 2 it can be proved that the J48 classification algorithm provides better result than the other algorithms by increasing the precision values. In this graph, methods are plotted in the x axis and the precision values are plotted in the y axis. The precision value is low by using kNN, NB and C4.5 algorithms for all enterprises. The precision value is significantly increased by using J48 algorithm for all enterprise databases. Thus the J48 algorithm is used to improve the performances prominently. From the experimental result, we can conclude that the J48 algorithm superior to other algorithms.

The values given in Figure 2 are illustrated as table format. The table 2 shows the different precision values for various algorithms and various enterprises. From the table values the J48 algorithm yields higher precision values compare than other algorithms. Hence the J48 is used to improve the solutions more accurately.

Table 2. Precision values

Enterprises	NB	kNN	J48	C4.5
Transactional database	0.75	0.69	0.89	0.8
Medical database	0.78	0.72	0.95	0.87
Agricultural database	0.79	0.75	0.97	0.91

Figure 2. Precision



### 3.3. Recall

Recall value is calculated based on the retrieval of information at true positive prediction, false negative. In healthcare data precision is calculated the percentage of positive results returned that are Recall in this context is also referred to as the True Positive Rate. Recall is the fraction of relevant instances that are retrieved,

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

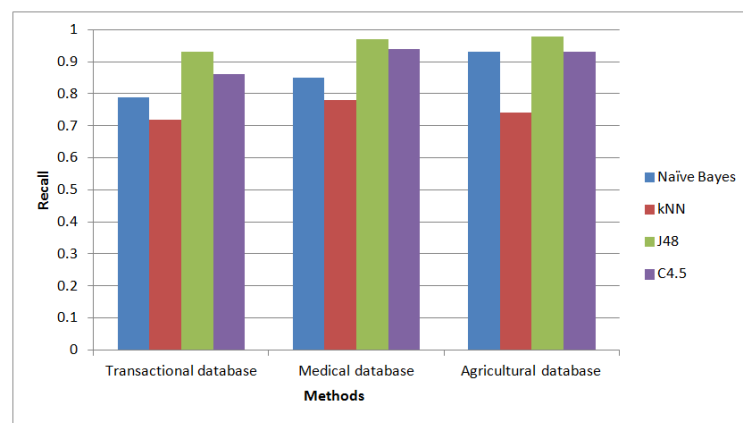
From the Figure 3 it can be proved that the J48 classification algorithm provides better result than the other algorithms by increasing the recall values. In this graph, methods are plotted in the x axis and the recall values are plotted in the y axis. The recall value is low by using kNN, NB and C4.5 algorithms for all enterprises. The recall value is significantly increased by using J48 algorithm for all enterprise databases. Thus the J48 algorithm is used to improve the performances prominently. From the experimental result, we can conclude that the J48 algorithm superior to other algorithms.

The values given in Figure 3 are illustrated as table format. The table 3 shows the different recall values for various algorithms and various enterprises. From the table values the J48 algorithm yields higher recall values compare than other algorithms. Hence the J48 is used to improve the solutions more accurately.

Table 3. Recall values

Enterprises	NB	kNN	J48	C4.5
Transactional database	0.79	0.72	0.93	0.86
Medical database	0.85	0.78	0.98	0.94
Agricultural database	0.93	0.74	0.97	0.93

Figure 3. Recall



### 4. Conclusion

The J48 algorithm improved the accuracy of transactional, medical and agricultural enterprise databases significantly. This research scenario is used to analyze and discover the manufacturing challenges as well as solutions by using the advanced data mining techniques. The performance of the system is improved higher in this scenario using J48 algorithm. In future we can extend the algorithms and it will lead the upcoming manufacturing products with more accurate results.

### 5. Acknowledgement

We the authors assure you that, this is our own work and also assure you there is no conflict of interest.

## 6. References

1. G.Geethakumari, Agrima Srivatsava. Big Data Analysis for Implementation of Enterprise Data Security, *IRACST-International Journal of Computer Science and Information Technology & Security (IJCSITS)*.2012; 2(4), 742-746.
2. Muhammad Shahbaz, Syed Athar Masood, Muhammad Shaheen, Ayaz Khan. Data mining methodology in perspective of manufacturing databases. *Journal of American Science*.2010; 6(11), 999-1012.
3. Vasnani, Pankaj Ramesh, Antim Arora, Nidhi Mittal. Data mining and OLAP Operations. *Journal of Global Research Computer Science & Technology (JGRCST)*. 2014; 1(3).
4. W.Craven Mark,W. Jude Shavlik. Using neural networks for data mining, *Future generation computer systems* . 1997; 13(2-3), 211-229.
5. Ben Jabeur Sami, F. A. H. M. I. Youssef. Predicting business failure using data mining methods. *Department of Research, Ipag Business School*, 2014.
6. K. Durgesh Srivastava, B. Lekha. Data classification using support vector machine, *Journal of Theoretical and Applied Information Technology*. 2010; 12(1), 1-6.
7. Chertov Oleg. Enterprise architecture model that enables to search for patterns of statistical information. *International Journal of Advanced Research in Artificial Intelligence*. 2013; 2(6), 1-5.
8. Chen Wei-Sen, Yin-Kuan Du. Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*. 2009; 36(2), 4075-4086.
9. Dhandayudam Prabha, Dr Illango Krishnamurthi. An improved Clustering Algorithm for customer segmentation. *International Journal of Engineering Science and Technology*. 2012; 4(2), 695-702.
10. Devi, Alias, P. Isakki, S. P. Rajagopalan. Analysis of Customer behavior using clustering and Association rules. *International Journal of Computer Applications*. 2012; 43(23), 19-26.
11. Clifton Phua, Vincent Lee, Kate Smith, Ross Gayler. A comprehensive survey of data mining-based fraud detection research, arXiv preprint arXiv:1009.6119. 2010.
12. Rygielski Chris, Jyun-Cheng Wang, C.David Yen. Data mining techniques for customer relationship management. *Technology in society*. 2002; 24(4), 483-502.
13. Laurent Probst, Erica Monfardini, Laurent Frideres, Steven Clarke, Dawit Demetri, Alain Kauffmann, PwC Luxembourg. Advanced manufacturing measurement technologies and robotics. Business Innovation Observatory. European Union. Sep 2013.
14. C. Sarada sowjanya, R.M. Srajan Ch. Application of data mining techniques for customer relationship management (CRM). *International Journal of Engineering Research & Technology (IJERT)*. 2013; 2(11).
15. K.Tawinunt, T.Phimonsathienand and Wannong Fongsuwan. A structural equation model of customer relationship management factors affecting customer retention of long stay travellers in thai tourism industry. *Research journal of business management*. 2015; 9(1), 1-24.
16. M.Duraij, V. Ranjani. Data mining applications in healthcare sectors a study. *International Journal of Scientific and Technology Research*. 2013; 2(10), 29-35.
17. D.Ramesh, B. Vishnu Vardhan. Data mining techniques and Applications to agricultural yield data. *International Journal of Advanced Research in Computer and Communication Engineering*. 2013; 2(9), 3477-3480.

The Publication fee is defrayed by Indian Society for Education and Environment (iSee). [www.iseeadyar.org](http://www.iseeadyar.org)

### Citation:

M.Srimathi, Dr. R.Umarani. Challenges and solutions for manufacturing products using advanced enterprise data mining approaches. *Indian Journal of Innovations and Developments*. 2015; 4 (5), September.