

# Application of Data Mining Techniques for Measuring and Predicting Employee Performances in Automotive Industry

**Pankaj Roy Gupta**

*Bharati Vidyapeeth (Deemed to be University), Pune, India (Research Scholar)*

**Dr Pravin Mane, Dr.Hema Mirji**

*Bharati Vidyapeeth (Deemed to be University), Pune, India (Assistant Professor)*

**Abstract:** The growth of integrated information technology based enterprise systems and social media platforms are helping automobile industries to generate large amount of databases and huge unstructured data on various business functions. Employees performance management monitoring and assessment in an automobile company is an important HR function, as it is directly related with the business results of the organization. Application of HR Analytics by using data mining techniques on employees and various productivity related data is used for extracting hidden insights for predicting employee performances.

Data Mining is a systematic process of discovering knowledge and useful information by extraction meaningful data patterns, profiles and trends using pattern recognition technologies such as classification, clustering, regression, artificial intelligence, neural networks, association rules, decision trees, machine learning, genetic algorithm, nearest neighbour algorithm etc. Data Mining tools are used for measuring the performances of the employees for predicting the success or failure of an organization.

The present applications of smart manufacturing practices using Industry 4.0 framework in automobile industry is generating large amount of data to be classified under Big Data category, which needs special data mining techniques for extracting intelligence from the data pattern.

This research contributes by exploring the data clustering, decision tree and other data mining techniques for evaluating and predicting the employee performances for automobile industry by considering various employee performance factors. Every automobile company has its own productivity and strength which depends upon the performance capability of its employees.

**Keywords:** Data Mining, Employee Performance, HR Analytics, Clustering, Decision Tree, Predictive Analytics, Big Data, Data Analytics.

## 1. INTRODUCTION

### 1.1 UNDERSTANDING DATA MINING

This paper proposes to understand the power of data mining techniques and their effective use for extracting intelligence and hidden knowledge on employee performances of the automotive industry's

10	ISSN 2576-5973 (online), Published by "Global Research Network LLC" under Volume: 6 Issue: 1 in Jan-2023 <a href="https://globalresearchnetwork.us/index.php/ajebm">https://globalresearchnetwork.us/index.php/ajebm</a>
	Copyright (c) 2023 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY). To view a copy of this license, visit <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>

various enterprise and non-enterprise sources of data. Data Mining is a systematic process of extracting and discovering useful hidden patterns and relationships from various data sources, i.e. structured and unstructured databases, texts and web data etc. using pattern recognition technologies such as statistical techniques, decision tree, neural networks, machine learning, artificial intelligence and genetic algorithms, [1].

In the process extracting hidden knowledge from various data sources, the data mining techniques, which are also known as Knowledge Discovery from Data (KDD) use the impure forms of noisy data, missing values, static, sparse, dynamic, relevance, interestingness, heterogeneity, algorithm efficiency, size and complexity, [2]. It is difficult to extract knowledge from the complex forms of data using pure statistical techniques, hence data mining techniques and tools are used for analyzing large datasets stored in files, databases, and other repositories.

The knowledge discovery process is an iterative sequence of the following steps, [2].

1. **Data cleaning** (remove noise and inconsistent data to make it pure for mining)
2. **Data integration** (multiple data sources may be combined)
3. **Data selection** (select datasets of different formats from the database and other data sources)
4. **Data transformation** (to make data ready for mining and various aggregation operations)
5. **Data mining** (process to extract data patterns)
6. **Pattern evaluation** (identify the data patterns based on measuring parameters for knowledge representation)
7. **Knowledge presentation** (visualization and knowledge representation for end-users)

The Knowledge Discovery in Databases (KDD) is a complex iterative process of discovering knowledge from various data sources. Data mining is one of the important activities in this entire process of extracting relevant knowledge for decision making.

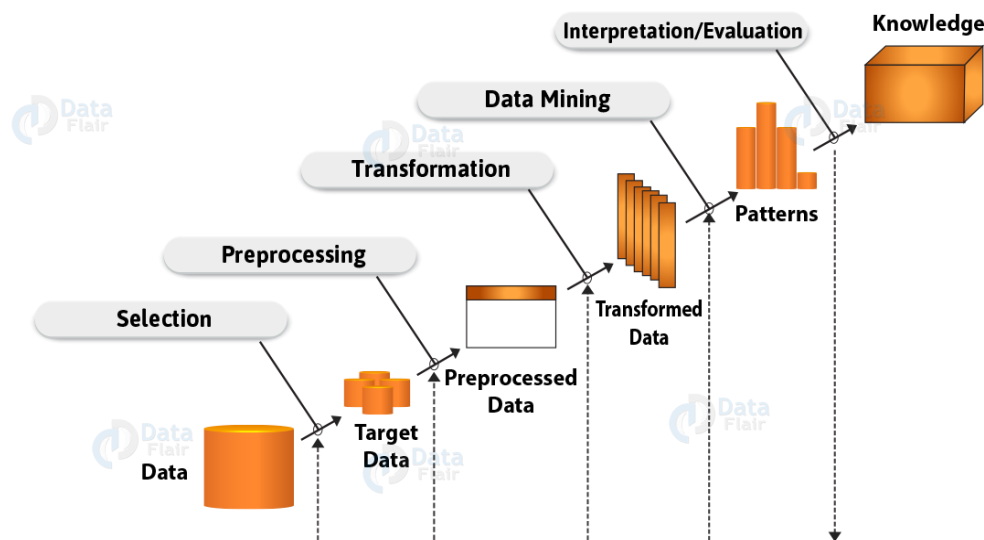


Fig.1: Knowledge Discovery Process Flow, [2]

Once the relevant and required knowledge is discovered and presented to the users, the evaluation measures can be enhanced, the mining can be further refined, new data can be selected or further transformed, or new data sources can be integrated, in order to get different, more appropriate results.

## 1.2 UNDERSTANDING EMPLOYEE PERFORMANCE ANALYSIS USING DATA MINING

The goal of this paper is to study the various data mining techniques used for measuring and predicting employee performances in automotive companies. Employee performance monitoring and assessment system in an organization helps employees to understand their behaviours toward organizational goals by allowing them to know what is expected from them, and it yields information for making employment related decisions, such as pay raises, promotion or releases, [55].

Organizations can improve the performances of their employees by way of monitoring the progress and predicting the expected outcomes as per the organizational goals and targets. Machine learning algorithms i.e. clustering algorithm and decision tree of data mining technique are used to find out the key characteristics of future prediction of an organization, [30]. Clustering is a method to group data into classes with identical characteristics in which the similarity of intra-class is maximized or minimized. This method is most widely used for future prediction, [35].

Employees are segregated into homogeneous groups according to their characteristics and abilities using clustering. Cluster analysis is used to segment employees in to groups according to their performance, [35]. Decision can be used for making meaningful decisions for the Employee. Based on the employee's performance results possible to take decision whether advanced training, talent enrichment or further qualification required or not. These applications also help administrative staff to enhance the quality of the organizations.

Employee performance evaluations are an important part of maintaining a motivated and skilled workforce. Every company maintains a confidential report form for measuring the quality of an employee throughout the year. The assessment is comparison with other stead of the same grade, should be recorded by putting their initials in the appropriate. For experiment confidential form has been used which are nineteen categories. These are Intelligence & Mental alertness, Initiative & Drive, Oral Expression, Written Expression, Ability to plan & organize work, Ability to supervise work, Quality of Work, Perseverance & Devotion to duty, Capacity to guide & train subordinates, Attitude towards superiors, Ability to work with others, Tact, Moral integrity, General sense of responsibility, Responsibility in financial matters, Personality, Public relations, Punctuality and Observance of security measures. The rating scale is the user input to the organization [4].

Performance Ratings	Performance Description
1	Poor
2	Average
3	Good
4	Very Good
5	Excellent

Data mining [5] consists a set of techniques that can be used to extract relevant and interesting

knowledge from data. It has several tasks such as association rule mining, classification and prediction and clustering. All classification methods are supervised learning techniques that classify data item into predefined class label. It is one of the most useful techniques in data mining to build classification models from an input data set. The used classification techniques commonly build prototypes that are used to guess future data developments.

12	ISSN 2576-5973 (online), Published by "Global Research Network LLC" under Volume: 6 Issue: 1 in Jan-2023 <a href="https://globalresearchnetwork.us/index.php/ajebm">https://globalresearchnetwork.us/index.php/ajebm</a>
	Copyright (c) 2023 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY). To view a copy of this license, <a href="https://creativecommons.org/licenses/by/4.0/">visit https://creativecommons.org/licenses/by/4.0/</a>

The basic algorithm for decision tree induction is a greedy algorithm that constructs decision trees in a top-down recursive divide- and-conquer manner [9]. Decision Tree generates a decision tree from the given training data. It is one of the most used techniques, since it creates the decision tree from the data given using simple equations depending mainly on calculation of the gain ratio, which gives automatically some sort of weights to attributes used, and the researcher can implicitly recognize the most effective attributes on the predicted target. As a result of this technique, a decision tree would be built with classification rules generated from it. Here Decision tree is used for predict Employee Performance.

### 1.3 EMPLOYEE PERPEFORMCE MEASUREMENT IN AUTOMOTIVE INDUSTRY

Automotive industry is continuously facing challenges in terms of its performance parameters due to continuously shifting market conditions, increase competitions, cost pressure and changing market scenarios, technology invasion, application of Industry 4.0 framework and various other global factors[8].

The following factors are impacting the automotive industry [9] and the employee performances.

1. Changing customer expectations
2. Disruptive impact of technology
3. Changing regulatory environments
4. Changing nature of mobility infrastructure
5. Globally interconnected industry

The automotive organizations have to manage and measure their employee performances due to the dynamic changes in the following areas:

1. Shifting from traditional manufacturing
2. Shorter cycle of innovation, development and production
3. Changing face of service and support system
4. Digital transformation
5. New jobs and associated skills
6. New skills and capabilities
7. Workforce composition and metrics

The employee performances in the automotive industry depends upon the role and areas of operations as the industry has a number of players from OEMs to tier 'n' suppliers, with each playing a critical role in the value chain. The industry also depends on a strong forward and backward linkages with other prominent industries such as logistics and transportation. This integration is set to take new shape as vehicles become more and more connected. Globally, the automotive Industry is already witnessing this phenomenon with many non- traditional partnerships (e.g. with IT product, OTT, electronics and telecom companies). This highly integrated value chain within the industry and its new interlinkages with allied and/or adjacent industries intensifies the impact of disruptive forces are making impact on the employee performances [15].

#### 1.3.1 Future Skills of employees for the automotive industry

The various jobs in the automotive sectors are linked with the skill sets of the employees working for this sector. The future skills required by the employees working with the automotive industry, which will make a very high impact on their performance parameters are as under, [56]:

13	ISSN 2576-5973 (online), Published by "Global Research Network LLC" under Volume: 6 Issue: 1 in Jan-2023 <a href="https://globalresearchnetwork.us/index.php/ajebm">https://globalresearchnetwork.us/index.php/ajebm</a>
	Copyright (c) 2023 Author (s). This is an open-access article distributed under the terms of Creative Commons Attribution License (CC BY).To view a copy of this license, visit <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>

1. Mechatronics
2. Artificial Intelligence (AI) and Machine Learning (ML)
3. Robotics
4. Digitization and digitalization
5. Industry 4.0
6. Data Analytics and Data Mining

## 2. Analysis on various data mining techniques used for employee performance measurements

Based on the literature review, many data mining techniques are used for extracting intelligence to analyse and predict employee performances in automotive industry. Some of the techniques used are as under:

### 2.1.1 Statistical Techniques

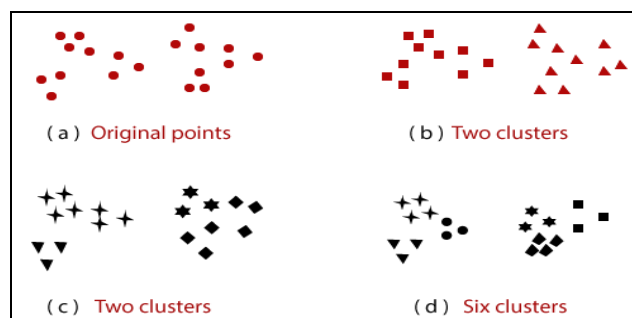
Use of statistical techniques are not data mining techniques but their use helps in determining patterns in datasets for building predictive models. There are different types of regressions in statistics but the simple form is Simple Linear Regression that contains one predictor and a prediction. The relations between the two can be mapped on a two dimensional space. In a multiple regression, more number of predictors can be used for solving a complex problem for making some better decisions.

### 2.1.2 CLUSTERING

Data mining techniques are classified as Supervised learning and Unsupervised learning. Clustering is a technique of unsupervised learning of grouping similar datasets into various groups of similar objects, [11]. Cluster analysis is used in analysing and predicting employee performances.

The following Clustering algorithms are used for predictive modelling [10].

1. Hierarchical clustering algorithms
2. K-means clustering algorithms
3. Density based clustering algorithms
4. Partition clustering algorithms
5. Special clustering algorithms
6. Grid based clustering algorithms

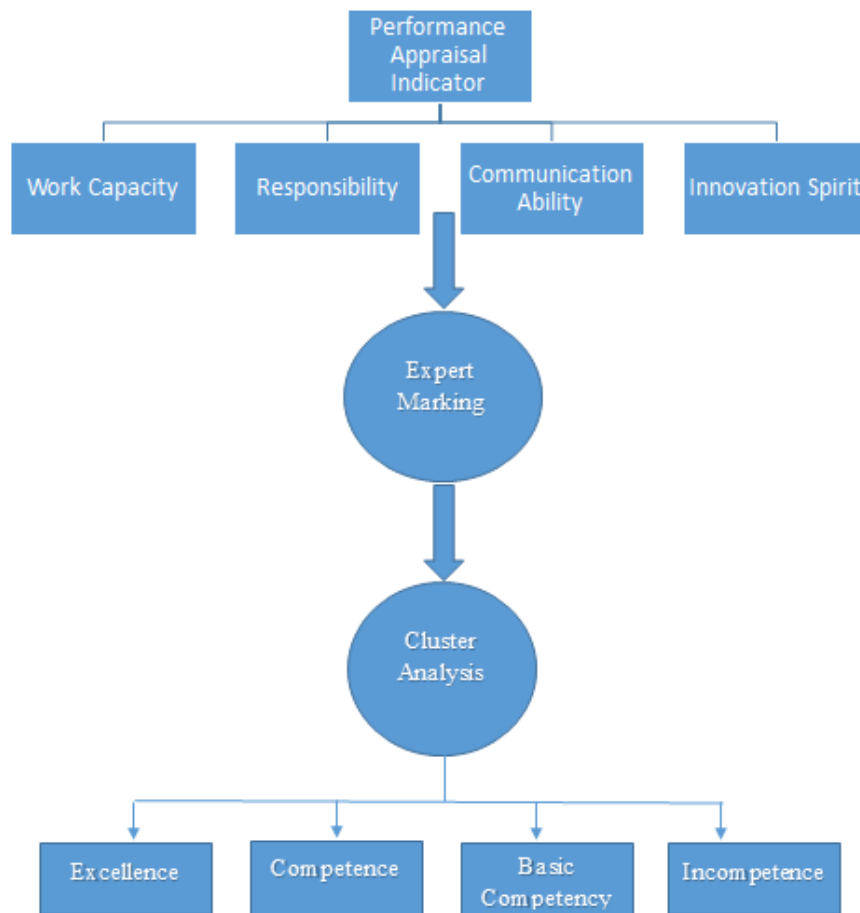


**Fig.2: Clustering Algorithm, [11]**

The K-means algorithms is used more often for employee performance measurement modelling, [12].

14	<p>ISSN 2576-5973 (online), Published by "Global Research Network LLC"  under Volume: 6 Issue: 1 in Jan-2023 <a href="https://globalresearchnetwork.us/index.php/ajebm">https://globalresearchnetwork.us/index.php/ajebm</a></p>
	<p>Copyright (c) 2023 Author (s). This is an open-access article distributed under the terms of  Creative Commons Attribution License (CC BY).To view a copy of this license,  visit <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a></p>

Employee are the key resources of the organization. The success or failure of an organization depends on the employee performance. Hybrid procedure based on Data Clustering and Decision Tree of Data mining method may be used to predict the employees' performance.



**Fig.3: Performance Appraisal Process Flow**

In Cluster Analysis, Employees can be evaluated on the basis on Rating Scale such as;

Very Good	5
Good	4
Average	3
Below Average	2
Poor	1

These Ratings are given in certain criteria as;

Cluster	Range of Category	No. of employee in different cluster			
		1 <sup>st</sup> Yr	2 <sup>nd</sup> Yr	3 <sup>rd</sup> yr	4 <sup>th</sup> yr
Excellent	Quality $\geq 4.5$	17	23	20	22
Good	$4.0 \leq \text{Quality} < 4.5$	20	22	25	26
Average	$3.0 \leq \text{Quality} < 4.0$	40	32	34	36
Poor	Quality $< 3.0$	23	23	21	16

### 2.1.3 Regression Analysis

Performance Evaluation can be done through Regression Analysis. Before starting the Regressing Analysis, the Independent Variables are identified.

In case of Performance Evaluation, the Independent variables can be; Work Condition, Pay and Promotion, Relationship with supervisor, Job Security, Fairness, Relationship with co-workers etc., we need to perform t-test.

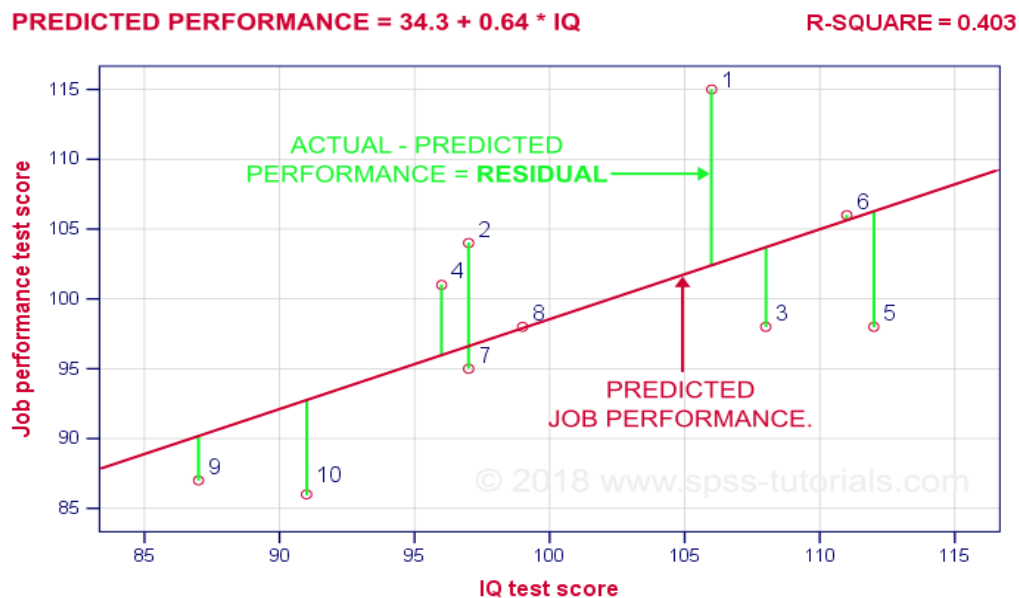


Fig.4: t-Test Regression Analysis on employee performance

### 3. CONCLUSIONS

This paper has reviewed various research papers published on Data Mining techniques, applications of data mining in measuring employee performance measurements and predictions, applications of data mining techniques in automotive industry for employee performance measurements and many related areas, i.e., artificial intelligence, Industry 4.0 framework and various other future technologies, which are going to impact on the employee performances in the automotive industry. Data Mining techniques are widely classified as classifications, regression, clustering, decision tree, artificial intelligence, machine learning etc., which are used for analysing employee performance data for extracting intelligence for employee performance predictions and decision making systems. There are many data mining tools which are used for different purposes during the various knowledge discovery stages in the data mining process.

This paper also indicates the capabilities of data mining techniques to provide effective employee performance measurements in automotive organizations. Many data mining techniques using various tools can be used to predict the employee performances. Extensive studies on the data mining techniques and their use in measuring and predicting employee performances, especially for automotive industry is carried away. The literature review highlights that there is a lack the studies in correlating the applications of data mining techniques on enterprise and non-enterprise employee performance data for measuring employee performance. The review leads to further scope in investigating the nonlinear cross-correlations between the various data mining techniques and their applications in measuring employee performance factors for automotive and other manufacturing industry.

#### 4. REFERENCES

1. O. Folorunso and A. O. Ogunde, "Data Mining as a Technique for Knowledge Management in Business Process Redesign," vol. 2, no. 1, pp. 33–44, 2004
2. O. R. Za, S. Pole, and C. Science, "Chapter I : Introduction to Data Mining," pp. 1–15, 1999
3. A. Sarker, S. M. Shamim, S. Zama, and M. Rahman, "Employee's Performance Analysis and Prediction using K-Means Clustering & Decision Tree Algorithm," vol. 18, no. 1, 2018.
4. Deloitte LLP, "Big data and analytics in the automotive industry Automotive analytics thought piece Contents," p. 16, 2015.
5. Lakshmi, T. M., Martin, A., Begum, R. M., & Venkatesan, V. P. "An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data". International Journal of Modern Education and Computer Science, 5(5), 18, 2013.
6. Al-Radaideh, Q. A., & Al Nagi, E., "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance". International Journal of Advanced Computer Science and Applications, 3(2), 2012.
7. Priyama, A., Abhijeeta, R. G., Ratheeb, A., & Srivastavab, S. "Comparative analysis of decision tree classification algorithms". International Journal of Current Engineering and Technology, 3(2), 334–337, 2013.
8. Deloitte LLP, "Big data and analytics in the automotive industry Automotive analytics thought piece Contents," p. 16, 2015.
9. Indian Automotive Sector, May 2019
10. S. Anitha Elavarasi and Dr J Akilandeswari and Dr. B Sathiyabhama, January 2011, A Survey on Partition Clustering Algorithms.
11. Survey of Clustering Data Mining Techniques, Pavel Berkhin, Accrue Software Inc, www.ijarcsee.com
12. S. Guha, R Rastogi and K Shim, 1998. CURE: An Efficient Clustering Algorithm for Lasrge Databases. Proc. ACM Int'l Conference Management of Data:73-84[6]
13. N. Nastasic, "Artificial Intelligence in HR: a No-brainer," *Pwc*, p. 8, 2017
14. N. Maan and D. Purwar, "Role of Artificial intelligence in MANET," *Adv. Res. Comput. Eng. ...*, vol. 5, no. 4, pp. 115–117, 2012.
15. D. Lin, C. K. M. Lee, H. Lau, and Y. Yang, "Strategic response to Industry 4.0: an empirical investigation on the Chinese automotive industry," *Ind. Manag. Data Syst.*, vol. 118, no. 3, pp. 589–605, 2018
16. PhridviRaj MSB., GuruRao CV (2013) Data mining – past, present and future – a typical survey on data streams. INTER-ENG Procedia Technology 12:255 – 263
17. Srivastava S (2014) Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering and Association Rule Mining. International Journal of Computer Applications (0975 – 8887) 88:..10
18. Soni N, Ganatra A (2012) Categorization of Several Clustering Algorithms from Different Perspective: A Review. IJARCSSE

19. Demšar J, Zupan B (2013) Orange: Data Mining Fruitful and Fun - A Historical Perspective. Informatica 37:55–60
20. Jain AK, Murty MN, Flynn PJ (1999) Data Clustering: A Review. ACM Computing Surveys, 31:264-323
21. Han J, Kamber M (2001) Data Mining. Kaufmann Publishers, Morgan
22. Rao IKR (2003) Data Mining and Clustering Techniques DRTC Workshop on Semantic Web, pp. 23-30
23. Mitra S, Pal KS, Mitra P (2002) Data Mining in Soft Computing Framework: A Survey. IEEE, 13: 3-14
24. Gupta GK (2012) Introduction to data mining with case studies PHI, New Delhi
25. Baker RID, Yacef K (2009) The State of Educational Data Mining:A Review and Future Visions. JEDM - Journal of Educational Data Mining, 1: 3-16
26. Kumar R, Kapil AK, Bhatia (2012) A Modified tree classification in data mining. Global Journals Inc. 12, 12: 58-63