

Identification of Contact Lens Types for Refractive Errors Using Iterative Dichotomiser3 (ID3) Algorithm

S. Sumitra¹, Dr. Ananthi Sheshasaayee²

¹Research Scholar, Bharathiar University, Coimbatore

² Associate Professor & Head, Department of Computer Science, Quaid – E – Millath Government College for Women, Autonomous, Anna Salai , Chennai 600002.

Abstract: *This paper throws light on identifying selective contact lens types for specific eye diseases from the list of available contact lenses in the current market. Refractive error of the eyes is a major problem in many countries. To measure the refractive errors many ophthalmologic methods such as retinoscope and phoropter were used. These instruments were used in checking the patient's eye for refractive errors. Information Technology (IT) has a tremendous growth in all disciplines. In the field of medical science, IT plays a vital role where the entire data is automated. In today's scenario, IT has a great impact in the field of health care informatics. ID3 algorithm is being used to identify the accurate contact lens for the refractive error. This phase of research provides an improvised method in decision-making about the refractive errors and identifying its corresponding contact lenses.*

Keywords: Contact less, Decision Tree, Refractive Errors, ID3 Algorithm

1. Introduction

In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm invented by Ross Quinlan used to generate a decision tree from a given dataset. Decision tree algorithms are widely implemented in the Classifier algorithms using recursive patterns with instance space. Decision trees are predictive model and data mining approaches that deal with decision trees bearing finite data. These predictive models are used in the area of statistics, data mining and machine learning techniques. Decision trees are also used as learning algorithms. The learning algorithms are created using the labeled trained classes. Decision trees resemble a flow chart where the leaf and non-leaf node are denoted along the attributes. The learning algorithms are classified as ID3, C4.5, Classification and Regression Tree (CART) and Chi-Squared Automatic Interaction Detector. Among the learning algorithms for this research, ID3 algorithm is considered.

2. Related Work

ID3 decision making algorithm has vast applications in various fields of science. This algorithm is used for the classification process. The algorithm aids in research areas like software engineering, forensic science, education systems, physical science, chemical science, web tools, and network security issues and mainly in the field of medicine. Classification algorithms provide the judgement for serious problem in any domain. ID3 algorithm was implemented in the field of clinical database to classify the amount of refractive error occurred among the school children with 187,513 samples. Classification was performed using entropy and information gain among the attributes and finally concluded that myopia disorder occurs in female students and astigmatism occurs among the male students[1]. Data mining methods are widely used in the field of computer forensics society to classify the inconstant, noisy and dispersive data. The research was performed using 100

samples with the modified ID3 algorithm. The result showed that ID3 possesses 8.9% of error rate, whereas the improved ID3 algorithm provides 5.4% error rate [2]

According to [3] in the process of knowledge processing during engineering, designing ID3 algorithm was used for determining injection molded parts. Using ID3, the key attributes are identified and implemented using fuzzy representation. The classification algorithm provided a decision tree which aided in analyzing the process. In the information security domain to identify the risk asset, the threat and the occurrence of vulnerability are identified using ID3 algorithm. Using the primary attributes like availability, confidentiality and integrity the algorithm was proposed and found a better way to identify the information assets. According to this classification the information security risk can be avoided [4]. ID3 algorithm plays a vital role in identifying the hidden classification rules among the mass students in the physical training and also in sports training. It aids the physical education teacher for planning the exercises opted to the students. The Physical Education teacher can also identify the sports and facility required by the candidate from this decision making system [5]. Logistics is the field where production and consumptions are executed. Logistics implements ID3 algorithm to evaluate the performance of the logistics process. The performance measures obtained from ID3 provide a successive production as a result[6]. According to [7], Data mining tools are high impacted in the field of education. By using ID3 and C4.5 algorithm prediction was performed to know the performance of the student for their future academics. Using this prediction method, it is judged how the student score marks in his future examinations.

3. Iterative Dichotomiser3 (ID3) Algorithm

ID3 is a simple decision learning algorithm developed by J. Ross Quinlan. It is a top down approach algorithm, which

works with the given training set of data. It tests the given data with its corresponding attributes in each and every node. The decision tree algorithm uses information gain for splitting the given data. The algorithm does not use any pruning technique and numeric attributes or any missing values. ID3 Algorithm uses two important steps to form the classification tree. They are Entropy and Information Gain.

Entropy: Entropy measures the information which refines the impurity of the arbitrary data collection. It measures the uncertainty of the data given in set S and is calculated as $H(S) = -\sum_{x \in X} p(x) \log_2 p(x)$

$$H(S) = -\sum_{x \in X} p(x) \log_2 p(x)$$

where S is used to calculate the entropy

X is set of Classes in S

p(x) is the proportion of the number of elements in class x to the number of elements in set S

If H(S) = 0 then classified.

Information Gain: Information gain is the measurement gained by finding the difference in entropy. By splitting the set S with the attribute A, the uncertainty in S is found. Hence the Information gain is calculated as

$$\text{Gain}(A, S) = H(S) - \sum_{t \in T} p(t) H(t)$$

Where H(S) is the set in entropy S

T is the subset from the splitting set S with the attribute A

t is the number of elements in t to the number of elements in set S

H(t) is the entropy of the subset t [1]

4. Proposed Work

4.1 ID3 Algorithm for Identifying Contact Lens Types using Refractive Vision Disorder Data Set.

ID3 was restricted to attributes with discrete values for not only the output variable but also the input variable.

Function ID3

Input: (R: A set of non target attributes)

C: The target attributes

S: A training set – returns a decision tree

Begin

If S is empty, return a single node with value Failure;

If S consists of all records with the same value of the target attributes, return a single leaf node with that value.

If R is empty, then return a single node with the value of the most frequent of the values of the target attribute that are found in records of S.

Let A be the attribute with largest Gain (A, S) among attributes in R;

Let $\{a_j, \forall j=1,2,\dots,m\}$ be the values of attributes A;

Let $\{S_j, \forall j=1, 2,\dots, m\}$ be the subsets of S consisting respectively, of records with value a_j for A;

Return a tree with root labeled A and arcs labeled a_1, a_2, \dots, a_m going respectively to the trees (ID3{R-A}, C, S1, ID3(R-{A}, C, S2),

ID3(R-{A}, C, S_m);

Recursively apply ID3 to subsets $\{S_j, \forall j=1,2,\dots,m\}$ Until they are empty

End

ID3 searches through the attributes of training instances and extracts the attributes the best separates the given examples. If the attribute perfectly classifies the training sets then ID3 stops; otherwise it recursively operates on the m (where m = number of possible values of an attribute) portioned subsets to get their best attribute. For each of the possible discrete values of this attribute a descendant node is created and the relevant training examples assigned to the correct node. The algorithm uses a greedy search to pick the best attribute and never looks back to consider earlier choices. The central focus of ID3 decision tree growing algorithm is to select the attribute to test at each node in the tree. For the selection of the attribute the concept of Entropy is used.

4.2 Refractive Error Vision Disorder Dataset Edifice

Vision disorder patient records are collected from various sources like optometrist, and opticians. Datasets are obtained by analysing the prescription provided by the eye doctor that have the information about the age of the patient, power in both left eye and right eye and the type of lens they can wear. Maximum of 126 patient records were accumulated. Among the stockpile, data were categorized according to the type of vision correction method the patient prefers. The 126 data were classified based on the category given in the table 1.1. The three major groups were categorized according to the preference of the patient. The groups were formed according to the preference in the eye glass, Contact lens and LASIK surgery. Table 1.1 was prepared based on the requirement of the patient. Among the 126 records 51 patients preferred eye glasses, 46 patients preferred contact lenses and 29 patients' selected LASIK surgery.

Table 1.1: Patient and their preferences for correcting methods

| No. of Patients | Eyeglass preference | Contact Lens preference | LASIK surgery preference |
|-----------------|---------------------|-------------------------|--------------------------|
| 126 | 51 | 46 | 29 |

Table 1.1 was again analyzed to get the necessary data for further processing. On analyzing the 51 eyeglass-preferred patients' records, the pediatric patients who were under 12 years of age totaled to 22 in number. These paediatrics were strictly advised to wear only eye glasses by the Optometrist/Ophthalmologist, since paediatrics will find difficult to handle the contact lenses. They were not advised to undergo LASIK surgery also. The remaining 29 patients under the eyeglass preference category were interested to undergo LASIK surgery. In the contact lens category, a total of 46 patients preferred contact lenses as the solution for refractive errors. Among the 126 record-set the 46 patients preferred wearing contact lenses and 4 patient's preference for refractive surgery were taken as a base source for the processing. Table 1.2 shows the patient and their preferences for contact lenses and refractive surgery.

Table 1.2: Patient and preferences to correct eye power

| No. of Patients | Preference of Contact Lens | Preference of Refractive Surgery |
|-----------------|----------------------------|----------------------------------|
| 50 | 46 | 4 |

Table 1.2 was considered as the primary data to find the corresponding contact lenses for the respective refractive

disorder. The table's data were quintessential that was required for classification. The patient records were formed based on their age and classified as Teen (Age 13 < 20), Adult (Age 20 < 40) and Old Age (40 and above) for the research. Further the dataset was categorized based on following seven parameters such as Simple Myopia, High Myopia, Hyperopia, High Hyperopia, Astigmatism, Blurred Vision and Cylindrical Correction

4.3 Conversion of Raw Data into Structured Data

Table 1.2 obtained from the optometrist, was further expanded to get the disorder formed in each patient's eye. The prescription given by the optometrist or ophthalmologist to patient contains the refractive error type on it. Those refractive errors are considered as the parameter for classification. From the survey obtained the following classifications have been obtained. Tables 1.3, 1.4 and 1.5 show the patient categories, namely, teen, adult and old age with their refractive error type and corresponding contact lens types respectively.

Table 1.3: Teenage Category with Corresponding Contact Lens Types

| S. No | Problem | Lens |
|-------|----------------|---------------------|
| 1 | Simple Myopia | RGP, Soft |
| 2 | Hyperopia | RGP, Soft |
| 3 | High Myopia | Soft |
| 4 | High Hyperopia | Not Applicable |
| 5 | Astigmatism | Spherical RGP, Soft |
| 6 | Presbyopia | Not Applicable |

Table 1.4: Adult Age Category with Corresponding Contact Lens Types

| S. No | Problem | Lens |
|-------|----------------|-----------------------------|
| 1 | Simple Myopia | Hybrid |
| 2 | Hyperopia | RGP |
| 3 | High Myopia | RGP, Soft, Hybrid |
| 4 | High Hyperopia | Soft, RGP, Hybrid |
| 5 | Astigmatism | Soft, Spherical RGP, Hybrid |
| 6 | Presbyopia | Not Applicable |

Table 1.5: Old Age Category with Corresponding Contact Lens Types

| S. No | Problem | Lens |
|-------|----------------|-----------------------|
| 1 | Simple Myopia | Not Applicable |
| 2 | Hyperopia | Not Applicable |
| 3 | High Myopia | Not Applicable |
| 4 | High Hyperopia | Not Applicable |
| 5 | Astigmatism | Not Applicable |
| 6 | Presbyopia | Bifocal Lens, Surgery |

The above three tables 1.3, 1.4 and 1.5 infer that High hyperopia and Presbyopia is not found among teens. According to the dataset Presbyopia is not found among adults and Simple myopia, hyperopia, high myopia, High hyperopia, astigmatism is not found among old age category

4.4 Implementation of ID3 for Identification of Contact Lens Types in Refractive Vision Disorder Dataset

The raw data set collected were categorized according to age and eventually subset of age group excluding paediatrics was selected, that is, age between 13 and 56 years. This data were

again mapped with the extracts obtained from an ophthalmologist, that is, with the lens used for the categories of a particular disease. To generate the decision tree entropy and information gain was calculated among the 50 record set.

Entropy and Gain Calculation

To design a decision tree using ID3 algorithm gain and entropy evaluation has been performed. The gain has been found in teen, adult and old age people with the number of occurrences of each refractive error type. The gain value with highest range will be the root node followed by the next smallest values.

$$\begin{aligned} \text{Gain}(S, \text{Age}) &= E(S) - 12/50 \quad E(\text{Teen}) - 26/50 \quad E(\text{Adult}) - 12/50 \\ E(\text{Old}) & \\ \text{Gain}(S, \text{Age}) & 0.95252 \\ \text{Gain}(S, \text{SM}) & 0.142348 \\ \text{Gain}(S, \text{Hyperopia}) & 0.128744 \\ \text{Gain}(S, \text{HM}) & 0.118 \\ \text{Gain}(S, \text{HH}) & 0.142348 \\ \text{Gain}(S, \text{A}) & 0.862608 \\ \text{Gain}(S, \text{BV}) & 0.799086 \end{aligned}$$

i. Entropy Calculation for Teen age groups

$$\begin{aligned} \text{Gain}(S_{\text{Teen}}, \text{SM}) & 0.642423 \\ \text{Gain}(S_{\text{Teen}}, \text{SH}) & 0.527593 \\ \text{Gain}(S_{\text{Teen}}, \text{HM}) & 0.515996 \\ \text{Gain}(S_{\text{Teen}}, \text{HH}) & 0.492413 \\ \text{Gain}(S_{\text{Teen}}, \text{A}) & 0.6910 \\ \text{Gain}(S_{\text{Teen}}, \text{BV}) & 0.492413 \\ \text{Gain}(S_{\text{Teen}}, \text{SM}) & 0.448642 \end{aligned}$$

ii. Entropy Calculation for Astigmatism with other refractive errors

$$\begin{aligned} \text{Gain}(S_{A_{\text{No}}}, \text{SH}) & 0.447004 \\ \text{Gain}(S_{A_{\text{No}}}, \text{HM}) & 0.457922165 \\ \text{Gain}(S_{A_{\text{No}}}, \text{HH}) & 0.447004 \\ \text{Gain}(S_{A_{\text{No}}}, \text{BV}) & 0.447004 \\ \text{Gain}(S_{A_{\text{No}}}, \text{SM}) & 0.448642 \end{aligned}$$

iii. Entropy Calculation for High Myopia with other refractive errors

$$\begin{aligned} \text{Gain}(S_{HM_{\text{No}}}, \text{SM}) & 0.548189 \\ \text{Gain}(S_{HM_{\text{No}}}, \text{SH}) & 0.5149 \\ \text{Gain}(S_{HM_{\text{No}}}, \text{HH}) & 0.48139 \\ \text{Gain}(S_{HM_{\text{No}}}, \text{BV}) & 0.48139 \end{aligned}$$

iv. Entropy Calculation for Simple Myopia with other refractive errors

$$\begin{aligned} \text{Gain}(S_{SM_{\text{No}}}, \text{SH}) & 0.4805258 \\ \text{Gain}(S_{SM_{\text{No}}}, \text{HH}) & 0.4539002 \\ \text{Gain}(S_{SM_{\text{No}}}, \text{BV}) & 0.4539002 \end{aligned}$$

v. Entropy Calculation for Simple Hyperopia with other refractive errors

$$\begin{aligned} \text{Gain}(S_{SH_{\text{No}}}, \text{HH}) & 0.53456 \\ \text{Gain}(S_{SH_{\text{No}}}, \text{BV}) & 0.53456 \end{aligned}$$

vi. Entropy Calculation for Old age groups

$$\begin{aligned} \text{Gain}(S_{\text{OLD}}, \text{BV}) & 0.496336 \\ \text{Gain}(S_{\text{OLD}}, \text{SM}) & 0.496336 \\ \text{Gain}(S_{\text{OLD}}, \text{SH}) & 0.496336 \\ \text{Gain}(S_{\text{OLD}}, \text{HH}) & 0.496336 \\ \text{Gain}(S_{\text{OLD}}, \text{HM}) & 0.496336 \end{aligned}$$

Gain (S_{OLD}, A) 0.496336

vii. Entropy Calculation for Adult group

- Gain (S_{ADULT}, SM) 0.133295
- Gain (S_{ADULT}, SH) 0.13295
- Gain (S_{ADULT}, HM) 0.17225
- Gain (S_{ADULT}, HH) 0.25477336
- Gain (S_{ADULT}, A) 0.47504
- Gain (S_{ADULT}, BV) 0.094582

viii. Entropy Calculation for Astigmatism with other refractive errors

- Gain (S_{ANo}, SM) 0.3000437
- Gain (S_{ANo}, SH) 0.2829508
- Gain (S_{ANo}, HM) 0.259098
- Gain (S_{ANo}, HH) 0.280386
- Gain (S_{ANo}, BV) 0.2572692

ix. Entropy Calculation for Simple Myopia with other refractive errors

- Gain (S_{SMNo}, SH) 0.550365
- Gain (S_{SMNo}, HM) 0.573705
- Gain (S_{SMNo}, HH) 0.657941
- Gain (S_{SMNo}, BV) 0.5000525

x. Entropy Calculation for High Hyperopia with other refractive errors

- Gain (S_{HHNo}, SH) 0.588009
- Gain (S_{HHNo}, HM) 0.684509
- Gain (S_{HHNo}, BV) 0.5288046

xi. Entropy Calculation for High Myopia with other refractive errors

- Gain (S_{HMNo}, SH) 0.585479
- Gain (S_{HMNo}, BV) 0.52

4.5 Root Node, Child Node Identification and Design

After calculating the entropy and information gain, the decision was formed. Based on the highest gain age is found to be the root node. Figure 1.1 shows the root node and its sub node.

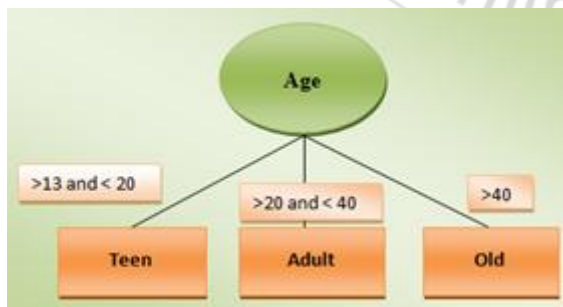


Figure 1.1: Patients Age Classification

Figure 1.1 shows the structure of the first level decision tree where the age is the root node followed by the category as teen, adult and old respectively. The tree structure was obtained based on the information gain obtained using ID3 algorithm process.

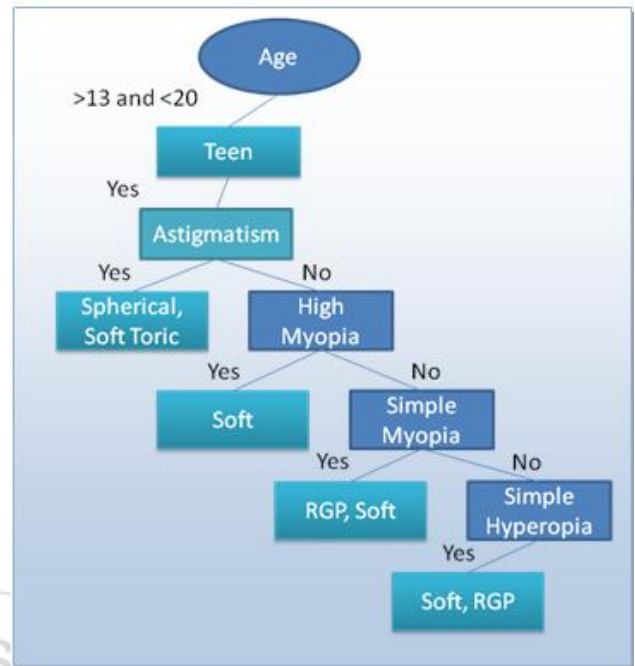


Figure 1.2: Hierarchy of Refractive Error with Lens Preferred in Teen Groups

Figure 1.2 shows the second level of the decision tree obtained for the teen age group. The gain calculated among the teen age group with their corresponding refractive errors. The result obtained from the information gain shows the hierarchy of the refractive error. Hence the refractive error and the used contact lenses are mapped in the decision tree.

Figure 1.3 shows the third level of the decision tree obtained for adult age groups. The gain calculated among the adult age group whose age ranges between 20 to 39 with their corresponding refractive errors are taken. The result obtained from the information gain shows the hierarchy of the refractive error. Hence the refractive error and the used contact lenses are mapped in the decision tree.

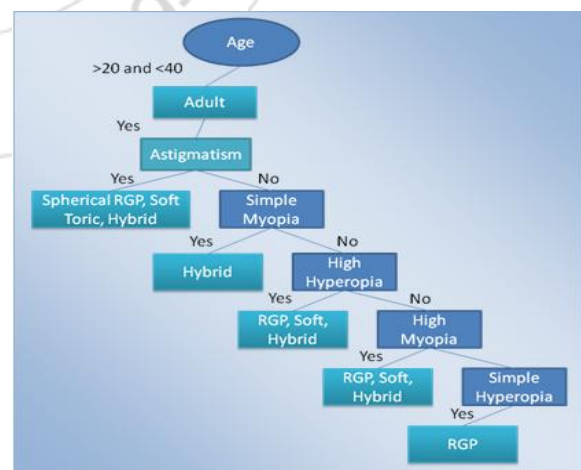


Figure 1.3: Hierarchy of Refractive Error with Lens Preferred in Adult Groups

Figure 1.4 shows the fourth level of the decision tree obtained for old age group. The gain calculated among the old age group with their corresponding refractive errors are taken.

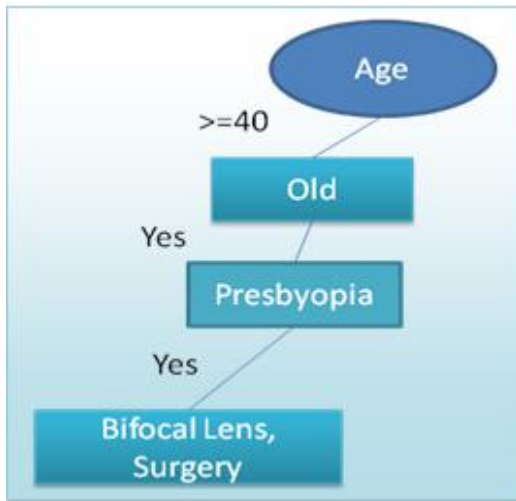


Figure 1.4: Hierarchy of Refractive Error with Lens Preferred in Old Age Groups

The result obtained from the information gain shows the hierarchy of the refractive error. The refractive error presbyopia occurs at this level. The reflections are that the old group people either use a bifocal lens or undergo refractive surgery according to their convenience. The decision tree, maps the refractive error and the used contact lenses.

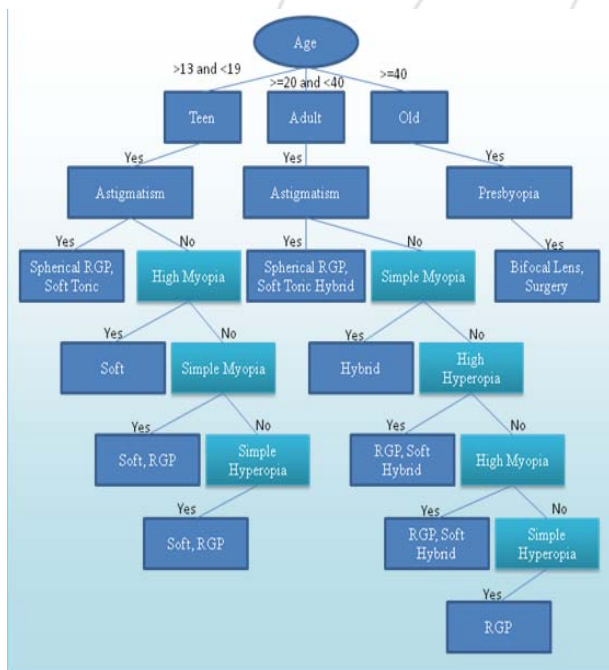


Figure 1.5: Decision Tree Refractive Errors and the Contact Lens Types

The decision tree in Figure 1.5 illustrates the hierarchical order of refractive errors and their corresponding contact lenses suitable for the age group people.

5. Conclusion

Various refractive errors occurring among the patients of different age groups were discussed. Based on the refractive errors the contact lenses were identified. By applying the process of ID3 algorithm appropriate decisions were framed. The Decision tree in Figure 1.5 illustrates the hierarchy of

refractive errors and appropriate correction methods to be selected. This decision tree can be used as a second opinion tool for the manufactures of contact lenses, practitioners of eye, opticians and patients.

References

- [1] Chandra Shekar D V, Sessa Srinivas V(2008), “Clinical Data Mining - An Approach for Identification of Refractive Errors”, Proceedings of the International MultiConference of Engineers and Computer Scientists 2008, (IMECS 2008), Vol. 1, 19-21.
- [2] Lu Qin (2010) , “Data Mining method based on computer forensics- based ID3 algorithm”, Published in The 2nd IEEE International Conference on Information Management and Engineering (ICIME), 2010.
- [3] Xinyu Shao, Guojun Zhang, Peigen Li, Yubao Chen (2001), “Application of ID3 algorithm in knowledge acquisition for tolerance design”, Journal of Materials Processing Technology (2001), Vol. 117, Issue 1-2, pp.66-74.
- [4] Hua Yong, Zhang Yunlong (2012), “Application of ID3 Algorithm in Information Asset Identification”, National Conference on Information Technology and Computer Science (CITCS 2012).
- [5] Quancheng Zhang, Kun You, Gang Ma (2011), “Application of ID3 Algorithm in Exercise Prescription”, International Conference on Electric and Electronics (EEIC 2011) in Nanchang, China on June 20–22, 2011, Vol. 3, pp. 669-675.
- [6] Yu Kui Sheng ,Wan Lian Lan (2011), “Research on Computer Integrated Manufacturing System Based on Integrated Logistics”, International Conference on Materials Engineering for Advanced Technologies (ICMEAT2011), 480-481:723-726, Jun 2011.
- [7] Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha, Vipul Honrao (2013), “Predicting Students’ Performance using ID3 and C4.5 Classification Algorithms”, International Journal of Data Mining & Knowledge Management Process (IJDMP), Vol. 3, No. 5, pp. 39-52.