Agropedology 2012, 22 (1), 35-42

An application of ID3 Decision Tree Algorithm in land capability classification

NIRMAL KUMAR, G. P. OBI REDDY, S. CHATTERJI, AND DIPAK SARKAR

National Bureau of Soil Survey and Land Use Planning, Nagpur – 440 033, India

Abstract : Land Capability Classification (LCC) is a technical classification in which the soil map units are grouped based on their characteristics for suitable use and management. LCC of a soil map unit is sought for, while planning its sustainable use and management and conservation practices. A pre-defined rule set for classifying a map unit would be of great help in developing decision support systems for land use planning of an area. Machine learning systems, which automatically learn rules from data, are often a very attractive alternative to manually constructing them. High speed, high precision and simple generating of rules by machine learning algorithms can be utilized to construct predefined rules for LCC of soil map units. The decision tree is one of the most popular classification algorithms currently in machine learning and data mining. Iterating Dichotomizer 3 (ID3) – A classical decision tree (DT) algorithm, was evaluated in land capability classification using data of 38 soil series of Wardha district, Maharashtra. Soil depth, slope, drainage, texture, erosion, and permeability were selected as attributes for land capability classification. A 10 - fold cross validation provided an accuracy of 86.84%. the results suggests that, explicit rules could be formulated with better accuracy for classifying complex soil-site data acquired over diversified land types.

Additional key words : *ID3, Decision tree, Entropy, Information gain, Machine learning*

Introduction

LCC provides information of the kind of soil, its location on the landscape, its extent, and its suitability for various uses which is needed for conservation planning, environmental quality, and generation of interpretive maps (Fenton 2005). The purpose of land capability classification systems is to study and record all data relevant to finding the combination of agricultural and conservation measures, which would permit the most intensive and appropriate agricultural use of the land without undue danger of soil degradation (Tripathi and Psychas 1992). USDA Land Capability Classification system (Klingebiel and Montgomery 1961) is undoubtedly the most used land classification system in the world (Rossiter 1994). LCC includes eight classes of which, first four are suitable for cropland and the limitations on their use and necessity of conservation measures and careful management increase from I through IV. The remaining four classes, V through VIII, are unsuitable for cropland, but may be used for pasture, range, woodland, grazing, wildlife, recreation, and esthetic purposes. Within the broad classes are subclasses, which signify special limitations such as (e) erosion, (w) excess wetness, (s) problems in the rooting zone, and (c) climatic limitations. Within the sub-classes are

the capability units which give some prediction of expected agricultural yields and indicate treatment needs. The capability units are groupings of soils that have common responses to pasture and crop plants under similar systems of farming.

The task of Land Capability classification occurs every time a soil surveyor identifies a map unit. A large and diversified dataset have already been created by previous surveys. A pre-defined rule set for automatically defining the LCC of the future map units being surveyed, will be of great help for developing decision support systems for land use planning and suggesting conservation and management practices. Machine learning algorithms build computer programmes that swift through databases automatically, seeking regularities or patterns. Strong patterns, if found, will likely generalize to make accurate predictions on future data. The DT is one of the most popular classification algorithms currently in machine learning and data mining (McQueen *et al.* 1995; Pal and Mather 2003; Lior *et al.* 2007; Gangrade *et al* 2009; Huang *et al.* 2010). DT approach employs tree-structured rules that recursively divide the data into increasingly homogeneous subsets based on splitting criteria (Safavian and Landgrebe 1991; deColstoun *et al.* 2003; Rogan *et al.* 2003; Trépos *et al.* 2012). The key to successful decision tree creation is in selecting the splitting criteria and the "best" attribute at each node (Rattray *et al.* 1999). The ID3 algorithm (Quinlan 1986), the most classical algorithm generating decision tree (Guan and Zeng 2011; Liu *et al.* 2011) uses entropy-based definition of information gain as splitting criteria (Quinlan 1986).

Recently, ID3 have been quite successfully used in various real-world applications. For instance, data related to biomedicine and food quality (Kinney and Murphy 1987; Li *et al.* 2008; Ture *et al.* 2009; Dębska and Swider, 2011; Angayarkanni and Banu 2012), pattern recognition (Liu 2008), Computational intelligence and security (Akhtar 2005; Ming and Shuxu 2009; Zhai and Liu 2010; Guan and Zeng 2011; Yanqin *et al.* 2011; Zhang *et al.* 2011; Zou *et al.* 2012), electronics (Du *et al.* 2011; Tan *et al.* 2011) and business and marketing (Xu 2005; Ke-wu *et al.* 2007; Chen 2011) to name some area of research. He *et al.* (2011) generated soil nutrient management zones using ID3 algorithm based on the contents of organic matter, total N, available P and available K in patch data of Dehui city, Jilin Province. Tamboli *et al.* (2012) evaluated ID3 DT for LCC with 12 simulated samples with soil depth, slope, and texture as attributes for LCC. The developed tree was not validated however. In the present study, we applied the same algorithm to real soil survey data in order to demonstrate the useful efficiency of the algorithm under practical circumstances. We interpret ID3 DT as the acquisition of structural descriptions from the training dataset of classified soil map units. The kind of descriptions found can be used for prediction, explanation, and understanding future soil map units.

Materials and Methods

ID3 was evaluated for LCC using data of 38 soil series (Sharma *et. al.* 2008) of Wardha district, Maharashtra. Soil depth, slope, drainage, texture, erosion, and permeability were selected as attributes for land capability classification (Table 1). The LCC were classified in accordance with Soil Survey Manual by All India Soil and Land use Survey Organization (AISLUS 1971). The LCC ranges from IIs to VIes.

Waikato Environment for Knowledge Analysis (WEKA) – an open source data mining tool – was used for ID3 algorithm. A brief of the ID3 algorithm is given below.

ID3

The ID3 algorithm a simple decision tree algorithm uses entropy-based definition of information gain as splitting criteria. Entropy characterizes the purity of any sample set. If the target attribute can take on v different values, then the entropy of set *(S)* relative to this v-wise classification is defined as

$$Entropy(S) = \sum_{i=1}^{v} -p_i \log_2 p_i \qquad (1)$$

where p_i is the proportion of S belonging to class i.

Table 1. Soil series description and attributes for LCC

Series	Depth	Slope	Erosion	Texture	Drainage	Permeability	Capability class
Kolona series	d5	d	e2	c	moderate	moderate	IIIes
Karanja series	d5	d	e2	c	moderate	moderate	IIIes
Nagjhari series	d5	d	e2	c	well	moderate	IIIes
Nijampur series	d5	b	e1	c	moderate	moderate	IIIs
Pachod series	d5	b	e2	c	moderate	moderate	IIIse
Vagholi series	d5	b	e2	c	moderate	moderate	IIIse
Thar series	d4	c	e2	c	moderate	moderate	IIIse
Anjangaon series	d4	c	e2	c	moderate	moderate	IIIse
Takli series	d5	c	e2	c	moderate	moderate	IIIse
Arvi series	d4	b	e2	gc	moderate	moderate	IIIse
Yakamba series	d4	b	e2	c	well	moderate	IIIse
Chamla series	d4	b	e1	c	moderate	moderate	IIs
Sirasgaon series	d3	c	e3	scl	well	moderate	IVes
Talani series	d2	c	e3	sl	well	moderate	IVes
Panthargavda series	d3	d	e3	gc	well	moderate	VIes
Parsodi series	d2	e	e3	scl	well	moderate	VIes
Hridi series	d4	c	e3	gc	well	rapid	IIIse
Chanakpur series	d3	b	e2	cl	well	rapid	IVs
Wadner series	d2	b	e2	gc	well	rapid	IVs
Pardi series	d2	b	e2	gc	well	rapid	IVs
Lakhandevi series	d3	e	e3	c	excessive	rapid	VIes
Mahakali series	d3	e	e3	cl	excessive	rapid	VIes
Karanii series	d3	h	e3	cl	excessive	rapid	VIes
Ashti series	d3	d	e3	gc	excessive	rapid	VIes
Kinala series	d2	e	e3	gl	excessive	rapid	VIes
Sewagram series	d2	d	e3	cl	well	rapid	VIes
Madni series	d2	d	e3	gc	well	rapid	VIes
Hewan series	d5	b	e1	c	moderate	slow	IIIs
Waigaon series	d4	b	e2	c	moderate	slow	IIIse
Karla series	d4	b	e2	c	moderate	slow	IIIse
Bothali series	d5	b	e2	c	moderate	slow	IIIse
Kondhali series	d5	b	e2	c	moderate	slow	IIIse
Wardha series	d5	b	e2	c	moderate	slow	IIIse
Lasanpur series	d5	b	e2	c	moderate	slow	IIIse
Malalpur series	d5	b	e2	c	moderate	slow	IIIse
Malakpur series	d5	c	e2	c	moderate	slow	IIIse
Sirpur series	d5	c	e2	c	moderate	slow	IIIse
Talegaon series	d5	b	e2	c	poor	slow	IIIse

Information gain is the expected reduction in entropy caused by splitting the training data set according to this attribute. More precisely, the information gain, *Gain(S, A)* of an attribute A, relative to a collection of examples S, is defined as

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

where *Values(A)* is the set of all possible values for attribute A, and S_v is the subset of S for which attribute

A has value v $(i.e., S_v = (S \in S | A(S) = v|))$.

The ID3 ceases to grow when all instances belong to a single value of a target feature or when best information gain is not greater than zero. ID3 does not apply any pruning procedure nor does it handle numeric attributes or missing values.

Performance evaluation

The classification performance of a model can be evaluated by the overall classification accuracy, the precision, the sensitivity (recall), F-measure, ROC area and the kappa value. In machine learning methods, such as the decision tree, the classification accuracy is often predicted by stratified 10-fold cross-validation (Weiss and Kulikowski 1991; Kohavi 1995; Kirchner *et. al.* 2006). In the process, original sample is randomly partitioned into 10 equal size sub-samples. One sub-sample is retained as the validation data for testing the model, and the remaining 9 sub-samples are used as training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 sub-samples used exactly once as the validation data. The 10 results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The classification accuracy assessment is calculated using the form of an error matrix (Congalton 1991). The error matrix is a square array and consists of the numbers of true positive (TP), false negative (FN), false positive (FP) and true negative (TN) classified examples.

Classification accuracy =

$$(TP + TN)/(TN + FP + FN + TP)*1000 \quad (3)$$

The Precision is the proportion of the examples which truly have class x among all those which were classified as class x. Precision may be calculated as:

$$\text{Precision} = TP/(TP + FP) \quad (4)$$

Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set. It is commonly also called sensitivity, and corresponds to the true positive rate. It is defined by the formula:

$$\text{Recall} = \text{Sensitivity} = TP/(TP + FN) \quad (5)$$

The *F-Measure* is a combined measure for precision and recall and calculated as

$$\text{F-measure} = 2*\text{Precision}*\text{Recall}/(\text{Precision} + \text{Recall}) \quad (6)$$

Receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the positives (TPR = true positive rate) *vs.* the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings. The kappa value is typically used as a measure of agreement between predicted and observed classes, and is calculated as

$$K = \frac{(TP + TN) - [((TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN))/N]}{N - [((TP + FN) \times (TP + FP) + (FP + TN) \times (FN + TN))/N]} \times 100 \quad (7)$$

Result and Discussion:

Induction of ID3 DT

Based on the data (Table 1), entropy of a root node containing the whole training set as its subset is calculated.

$$\text{Initial Entropy (LCC)} = -\tfrac{3}{38}\log_2\tfrac{3}{38} - \tfrac{2}{38}\log_2\tfrac{2}{38} - \tfrac{18}{38}\log_2\tfrac{18}{38} - \tfrac{1}{38}\log_2\tfrac{1}{38} - \tfrac{1}{38}\log_2\tfrac{2}{38} - \tfrac{3}{38}\log_2\tfrac{3}{38} - \tfrac{9}{38}\log_2\tfrac{9}{38} \quad 2.166\,(P)$$

To identify the attribute to start the decision tree, information gain at each of the attribute is shown in table 2. Since the attribute *slope* is having maximum information gain, the root node will start from *slope*.

Table 2. Gain at the first node

Attribute	Gain
Slope	1.1487
Erosion	1.1061
Depth	1.0813
Texture	0.8657
Drainage	0.848
Permeability	0.7434

In case of attribute *slope*, the entropies (Table 3) suggest that *slope b, c,* and *d* need to be split again,

Table 3. Entropies of classes of attribute slope

Slope	Entropy
b	1.404678
c	0.811278
d	0.985228
e	0
g	0

Table 4. Information gain at child nodes of attribute slope

Attributes	Slope b	Slope c	Slope d
Depth	0.803	0.811	0.985
Permeability	0.736	0.811	0.522
Erosion	0.65	0.467	0.985
Texture	0.516	0.811	0.985
Drainage	0.508	0.467	0.522

Table 5. Entropies of classes of attribute depth under different slope classes

Depth	Slope b	Slope c	Slope d
d2	0	0	0
d3	0	0	0
d4	0.722	0	0
d5	0.722	0	0

Table 6. Information gain at child nodes of attribute depth

Attribute	Depth d4	Depth d5
Erosion	0.7219	0.7219
Permeability	0.171	0.0341
Drainage	0.0729	0.0323
Texture	0.0729	0

whereas, since entropy at *slope e* and *g* are zero, these are terminated. ID3 was again applied to each child node of this root. At all the three nodes, the information gains (Table 4) for attribute *depth* is found to be highest. Thus the next splitting criteria will be *depth* for all these child nodes of attribute *slope*. Again entropies at each child node at next step (Table 5) indicates that only the child nodes *d4* and *d5* of *slope* are needed to be split as others have entropy zero. Based on the information gain values (Table 6), the attributes to split these child nodes are *erosion* in both the cases.

```
Slope = b
|   Depth = d2: IVs
|   Depth = d3: IVs
|   Depth = d4
|   |   Erosion = e2: IIIse
|   |   Erosion = e3: null
|   |   Erosion = e1: IIs
|   Depth = d5
|   |   Erosion = e2: IIIse
|   |   Erosion = e3: null
|   |   Erosion = e1: IIIs
Slope = c
|   Depth = d2: IVes
|   Depth = d3: IVes
|   Depth = d4: IIIse
|   Depth = d5: IIIse
Slope = d
|   Depth = d2: VIes
|   Depth = d3: VIes
|   Depth = d4: null
|   Depth = d5: IIIes
Slope = e: VIes
```

Fig. 1. Decision tree with ID3 algorithm

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g   <-- classified as
  2  0  0  0  0  0  0 |  a = IVs
  0  1  0  0  0  0  0 |  b = IVes
  0  0  9  0  0  0  0 |  c = VIes
  0  0  0  0  1  0  0 |  d = IIs
  0  1  0  0 17  0  0 |  e = IIIse
  0  0  0  0  1  1  0 |  f = IIIs
  0  0  0  0  0  0  3 |  g = IIIes
```

Fig. 2. Confusion Matrix with ID3 algorithm

Performance of DT

A 10 – fold cross validation was applied the developed model by ID3. Out of 38 samples, 33 were classified correctly and 3 incorrectly, whereas,

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	0.833	IVs
	1	0.029	0.5	1	0.667	0.736	IVes
	1	0	1	1	1	1	VIes
	0	0	0	0	0	0.5	IIs
	0.944	0.111	0.895	0.944	0.919	0.922	IIIse
	0.5	0	1	0.5	0.667	0.75	IIIs
	1	0	1	1	1	1	IIIes
Weighted Avg.	0.917	0.056	0.906	0.917	0.904	0.917	

Fig. 3. Class-wise precision, recall, F-measure and ROC area

remaining 2 were unclassified. The overall accuracy was 86.84% with a kappa coefficient of 0.87. The average precision, recall and ROC area were 0.906, 0.917, and 0.917, respectively. The final tree developed by ID3 algorithm and the confusion matrix are shown in figure 1 and 2, respectively. The class-wise precision, recall, F-measure and ROC area is shown in figure 3.

Conclusion

Machine learning present the basic theory of automatically extracting models from data, and then validating those models. Here we have focused on ID3 DT, in which classification results from a sequence of logical steps. These are capable of representing the most complex problem given sufficient data. Fundamental to DT is selection of the "best feature" for splitting the data. Feature selection, by identifying the most salient features for learning, focuses a learning algorithm on those aspects of the data most useful for analysis and future prediction. The splitting criterion for ID3 is entropy gain. Slope is found to be the best attribute at the starting node in ID3 algorithm owing to have maximum entropy gain. In a 10-fold cross validation ID3 gives 86.84% accuracy. The results indicate that decision tree algorithms have good potential in land capability classification of soil survey data. Explicit rules could be formulated with better accuracy for classifying complex soil-site data acquired over diversified land types.

References

AISLUS 1971.All India Soil and Land Use Survey, Soil Survey Manual, Indian Agricultural Research Institute (IARI) Publ. New Delhi.

Akhtar, S., 2005. A proposed model to use ID3 algorithm in the classifier of a network intrusion detection system. 9th International Multitopic Conference, IEEE INMIC, 1 – 8.

Angayarkanni, P.S., Banu, K.N., 2012. MRI mammogram image classification using ID3algorithm. IET Conference on Image Processing, 1 – 5.

Chen, C., 2011. The apply of ID3 in stock analysis. 6th International Conference on Computer Science & Education (ICCSE), 24 – 27.

Debska, B., Swider, B.G., 2011. Decision trees in selection of featured determined food quality. *Analytica Chimica Acta* **705**, 261– 271.

deColstoun, E.C.B, Story, M.H., Thompson, C., Commisso, K., Smith, T.G., Irons J.R., 2003. National Park vegetation mapping using multi-temporal Landsat 7 data and a decision tree classifier. *Remote Sensing of Environment* **85**, 316–327.

Du, J., Cui, H., Li, W., Zhao, Z., 2011. Fault diagnosis of vacuum circuit breakers based onID3 method. 1st International Conference on Electric Power

Equipment - Switching Technology (ICEPE-ST), 283 – 286.

Fenton, T.E., 2005. Land Capability Classification. In Encyclopedia of Soil Science, Second Edition. CRC press. Pp 962-964.

Gangrade, A., Patel, R., 2009. Building privacy-preserving C4.5 decision tree classifier on multiparties. International *Journal of Computer Science and Engineering* **1**, 199-205.

Guan, C., Zeng, X., 2011. An improved ID3 based on weighted modified information gain. Seventh International Conference on Computational Intelligence and Security, 1283 – 1285.

He, L., Liying, C., Guifen, C., Dexin, L., 2011. Delineating soil nutrient management zones based on ID3 algorithm. International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), 1155 – 1159.

Huang, Y., Lan, Y., Thomson, S.J., Fang, A., Hoffmann, W. C., Lacey, R.E., 2010. Development of soft computing and applications in agricultural and biological engineering. *Computers and Electronics in Agriculture* **71**, 107–127.

Ke-wu, Y., Jin-fu, Z., Qiang, S., 2007. The application of ID3 algorithm in aviation marketing. IEEE International Conference on Grey Systems and Intelligent Services, GSIS. 1284 – 1288.

Kinney, E.L., Murphy, D.D., 1987. Comparison of the ID3 algorithm versus discriminant analysis for performing feature selection. *Computers and Biomedical Research* **20**, 467-476.

Kirchner K., T¨olle, K.H., Krieter, J., 2006. Optimisation of the decision tree technique applied to simulated sow herd datasets. *Computers and Electronics in Agriculture* **50**, 15–24.

Klingebiel, A.A., and Montgomery, P.H., 1961. Land capability classification. Agriculture handbook no 210. Soil conservation service, Washington D.C. US Department of Agriculture (USDA).

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada. Morgan Kaufmann, San Francisco, USA.

Li, Z., He, B., Chen, X., 2008. Analysis and disposal of pathogenic factors of exogenous affections based on the ID3 decision tree method. 7th World Congress on Intelligent Control and Automation (WCICA), 7509 – 7514.

Lior, R., Maimon, O., Maimon, O. Z., 2007. Data Mining With Decision Trees: Theory and Applications. Series in Machine Perception and Artificial Intelligence 69. World Scientific. Israel. 264pp.

Liu, X., Wang, D., Jiang, L., Chen, F., Gan, S., 2011.A novel method for inducing ID3 decision trees based on variable precision rough set. Seventh International Conference on Natural Computation (ICNC) 1, 494 – 497.

Liu, Y., Zhang, D., Lu, G., 2008. Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition* **41**, 2554-2570.

McQueen, R.J., Garner, S.R., Nevill-Manning, C.G., Witten, I.H., 1995. Applying machine learning to agricultural data. *Computers and Electronics in Agriculture* **12**, 275–293.

Ming, Y., Shuxu, G., 2009. Research and realization of security policy in IPSec based on ID3 algorithm. International Conference on Multimedia Information Networking and Security 2, 518 – 521.

Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* **86**, 554 – 565.

Quinlan, J. R., 1986. Induction of decision trees. Machine Learning **1**, 81-106.

Rattray, J., Floros, J.D., Linton, R.H., 1999. Computer-aided microbial identification using decision trees. *Food Control* **10**, 107-116.

Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., Fischer, C., 2003. Land-cover change monitoring with classification trees using landsat TM and ancillary data. *Photogrammetric Engineering & Remote Sensing* **69**, 793–804.

Rossiter, D.G., 1994. Land Evaluation Course Notes Part 7: Non-FAO Land Classification Methods. Pp30.

Safavian, S. R., Landgrebe, D., 1991. A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics 21, 660–674.

Sharma, J.P., Mandal, C., Raja, P., Nair, K.M., Bhaskar, B.P., Sarkar, D., 2008. Reconnaissance soil survey, mapping, correlation and classification of Wardha district, Maharashtra. NBSS Publication 595, NBSS&LUP, Nagpur. pp. 126.

Tamboli, N.M., Kamble, A.M., Metkewar, P.S., 2012. LCC Decision tree analysis using ID3 *International Journal of Computer Application* **41**, 19-22.

Tan, Y., Qi, Z., Wang, J., 2011. Applications of ID3 algorithms in computer crime forensics. International Conference on Multimedia Technology (ICMT), 4854 – 4857.

Trépos, R., Masson, V., Cordier, M.O., Chantal, G.O., Jordy, S.M., 2012. Mining simulation data by rule induction to determine critical source areas of stream water pollution by herbicides. *Computers and Electronics in Agriculture* **86**, 75–88.

Tripathi, B.R., Psychas, P.J., 1992. The AFNETA alley farming training manual - Volume 2: Source book for alley farming research pp 227.

Ture, M., Tokatli, F., Kurt, I., 2009. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications* **36**, 2017-2026.

Weiss, S. M., Kulikowski, C. A., 1991. Computer systems that learn. San Mateo, CA: Morgan Kaufman Publishers.

Xu, B., 2005. Managing customer satisfaction in maintenance of software product family via ID3. Proceedings of international conference on machine learning and cybernetics 3, 1820 – 1824.

Yanqin, Z., Peide, Q., Yuemei, H., 2006. Design and Optimization of VPN Security Gateway Communications and Networking in China. First International Conference on China Com., 1- 4.

Yiwen Zhang; Lili Ding; Yun Wang. Research and design of ID3 algorithm rules-based anti-spam email filtering. IEEE 2nd International Conference on Software Engineering and Service Science (ICSESS), 572 – 575.

Zou, K., Sun, W., Yu, H., Liu, F., 2012. ID3 Decision Tree in Fraud Detection Application. International Conference on Computer Science and Electronics Engineering (ICCSEE) 3, 399 – 402.

Received : April 2012 *Accepted : June 2012*