iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75194

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Efficient Decision Tree and Outlier Analysis on Neural and Financial Data Towards Detection of Money Laundering

J. Stanly Thomas[1], Dr N Rajkumar[2]
[1]*Research Scholar, Dravidian University, India,*
[1]*Director (Research), KGISL Institute of Technology, India*

Abstract: The adaptation of new technologies in the banking industry is continuous and growing drastically to replace the counter transactions in the banking. On the other hand, fraudulent transactions are hindering reputation and profitability aspects of this industry. In order to prevent this deter, the real time genetic and analytical tool is required for the same. This laid the corner stone for this research work, which is built with suitable algorithm to analyze each customer by their pattern on transactions to avoid money laundering in the bank account. The real challenging task under this research work is to classify and cluster all the transactions and customer base which is exceptionally very large database. The taste of the prevention is fully depends on the above say as the filtration of necessary data increase the accuracy of this research work. The Decision Tree Classification algorithm is constructed as a basement for this research work. Each of the balanced decision trees are enabled with weighed average which identifies the risk factor and cluster index. This research work is loaded with total of thirty indicative bullets under the decision tree and further clustered with five groups. Based on the outcome of this decision tree and its loaded weight, Data Cube outlier analysis shall find the relevance of the same which shall cause money laundering in near future.

Keywords: Decision Tree Classification, Cluster Analysis, Association rules, Genetic Algorithms, Data Cube Outlier Analysis,

I. INTRODUCTION

The approach of handling exceptionally larger database is started with data cleaning and listing of indicative bullets which shall resulted in money laundering. This research work is fully loaded with database such as customer base, transaction base, risk base in addition to outside agency comments. The following indicative alert tabulation is identified to start with the decision tree operation.

II. INDICATIVE ALERT INDICATORS

The approach of handling exceptionally larger database is started with data cleaning and listing of indicative bullets which shall resulted in money laundering. This research work is fully loaded with database such as customer base, transaction base and risk base in addition to outside agency comments. The following indicative alert tabulation is identified to start with the decision tree operation.

Sl.No.	Alert Indicator and Identity	Indicative Rules/ Scenario	
1	Un successful customer on-boarding (D_1)	Customer not on-boarded once the requirement of Know Your Customer (KYC) advised.	CUSTOMER
2	Preliminary onboarding undertaken with counterfeit documents (D_2)	Customer undertaken preliminary on-boarded with counterfeit documents or morphed evidences.	
3	KYC Document not able to validate (D_3)	Authenticity of the documents presented for identification not verifiable example document issued by foreign entities	

4	Non existence of furnished address (D_4)	Address mentioned in the address proof (KYC document) is not present	BASE
5	Wrong address identified (D_5)	Identified that Address furnished by customer is found to be wrong during scrutinizing the documents	
6	Beneficiary genuine not established (D_6)	Cumbersome legal structures and due to this beneficial ownership not established.	
7	Established criminal records found against customer (D_7)	Criminal offences observed against the customer and customer has been under the offence processing from law enforcement agencies.	RISK BASE
8	Terrorist activities and Terrorist finance offences against customer. (D_8)	Customer subject to investigation on the offences under Terrorist activities or Terrorist Finance.	
9	Media report established against customer under critical criminal background (D_9)	Clear matching of customer details with media report on criminal offences	
10	Media report established against customer under Terrorist or Terrorist finance (D_{10})	Clear matching of customer details with media report on terrorist or terrorist finance.	
11	Broken customer transaction (D_{11})	After raising of queries of source of fund customer broke the transactions and not initiated transaction	TRANSACTION & RISK BASE TRA
12	Tense behavior of customer (D_{12})	Customer clearly establish tense and panic	
13	Customer is vigilant (D_{13})	Customer is over careful and vigilant while establishing source of fund or transaction genuine	
14	Customer furnishes erratic or uneven information (D_{14})	• Alters the information once raised queries on the data provided by customer. • Customer caution full provided minimum information and provided such information which cannot be established	
15	Customer acts as a Third party power of attorney holder (D_{15})	• No knowledge on transaction and money put through over the transactions. • Reaches unknown party for undertaking each transactions • Customer appears to be with unrelated or unknown representatives during undertaking of transactions.	
16	Chain of individual customers formed as unkown group (D_{16})	Individual money laundering and similar transactions are being undertaken on group of individual customers. During undertaking of such transactions all the individual customers are present at the same time however pretend to behave as an unknown individuals.	
17	Customer choosing farthest branches (D_{17})	Customer chosen to undertake CBS transactions through farthest branches of the same Bank.	
18	Customer furnishes various	To avoid connection between transactions	

	identifications on various occasions (D_{18})	customer deliberately establishes with different identification at each transactions.	NSACTION & RISK BASE
19	Very caution of reporting of transactions (D_{19})	Customer not willing to report his/her transactions to various regulatory requirements and enquiring of the same as well as making effort to avoid the reporting.	
20	Source of fund not established (D_{20})	Customer unable to establish clear source of fund and provides inconsistent information every time.	
21	Cumbersome transaction flow (D_{21})	Transaction flow is very cumbersome for its purpose which seems intentional.	
22	Transaction rationale is not related to customer (D_{22})	The transaction value and periodicity relationship with transaction is not rationale and not related to customer.	
23	Transaction inconsistent with business (D_{23})	Nature of the transaction is not related to the background of customer or his/her business.	
24	unauthorised Non profitable organization transactions (D_{24})	Non profitable organization received Foreign inward remittance not authorized or not approved by regulator	
25	Cross- border Inward remittance into Non Resident External (NRE) account (D_{25})	Source of the remittance is unknown	
26	Suspicious cross-border remittance (D_{26})	Cross border Remitter details are changing every time	
27	Continuous cross-border remittance decline (D_{27})	Cross border remittance decline continuously due to non clear information on remitter/remitter who is involved in suspicious/criminal activities	
28	Public grievances received against customer (D_{28})	Public grievances received on the account of undertaking fraudulent activities etc.	AGENCY
29	Agent triggered alert (D_{29})	Agent raised suspicion alert over the account/customer/transactions.	
30	Institution triggered Alert (D_{30})	Other institutions/subsidiaries raised suspicion alert over the account/customer/transactions.	

TABLE 1 ALERT INDICATOR INPUTS

The above tabulation is derived from a banking industry to analyse the symptoms about the Money Laundering on every stage to prevent the same. The entire levels are classified into five clusters as mentioned earlier to automate the process in full phase and resulted in instant exploration of the money laundering. Each cluster is identified by its unique notation such as C_1,C_2,C_3,C_4 and C_5 and the relevant plotted areas are similarly classified as P_1, P_2, P_3, P_4 and P_5 respectively.

III. HYPERPLANES ON DECISION TREE CONSTRUCTION

All the points denoted by "a" are defined as the hyperplane hy(a) if satisfies the below derivations.

$$hy(a) : w^T a + b = 0$$

The weight vector W and its offset of the hyperplane "b" from the original origin. Decision tree algorithm intakes the weight vector which parallel to any of the dimensions or axis *aj*. The weight vector W are one allowed under the vectors $\{v_1, v_2, \ldots, v_d\}$ and relation model for decision tree algorithm defines the following points of hyperplanes.

$$hy(a) : v_j^T a + b = 0, \text{ which implies that}$$
$$hy(a) : a_j + b = 0$$

The choose features of offset "b" yield with various hyperplanes parallel to dimensions a_j .

IV. DATA CUBE OUTLIER ANALYSIS

Based on the decision tree rules and classification, the outlier range from the sample data is plotted in Cube format so that it shall be easily identified from the visualization tool. The number of attributes involved in data mining is very large and this can be in millions. The knowledge amongst the very large dimensional entity or hyperspace is very significant since hyperspace is

Consider the n x d data matrix

$$\mathbf{D} = \left(\begin{array}{c|cccc} & X_1 & X_2 & \cdots & X_d \\ \hline \mathbf{x}_1 & x_{11} & x_{12} & \cdots & x_{1d} \\ \mathbf{x}_2 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n & x_{n1} & x_{n2} & \cdots & x_{nd} \end{array}\right)$$

where each point $x_i \in R^d$ and each attribute $x_i \in R^n$.

The minimum and maximum values for each attribute X_j be given as

$$\min(X_j) = \min_i \{x_{ij}\} \qquad \max(X_j) = \max_i \{x_{ij}\}$$

The data cube can be considered as a d- dimensional hyper-rectangle, defined as

$$R_d = \prod_{j=1}^{d} \left[\min(X_j), \max(X_j)\right]$$
$$= \left\{\mathbf{x} = (x_1, x_2, \cdots, x_d)^T \mid x_j \in [\min(X_j), \max(X_j)], \text{for } j = 1, \ldots, d\right\}$$

cumbersome and graphs are highly critical under multi dimensions.

Assume the data is centered to have mean $\mu = 0$. Let m denote the largest absolute value in D, given as

$$m = \max_{j=1}^{d} \max_{i=1}^{n} \left\{|x_{ij}|\right\}$$

The data hyperspace can be represented as a data cube, centered at 0 , with all sides length $l=2m$, given as

$$H_d(l) = \left\{\mathbf{x} = (x_1, x_2, \cdots, x_d)^T \mid \forall i,\ x_i \in [-l/2, l/2]\right\}$$

The data cube is one dimension, $H_1(l)$, represents as interval, that in two dimensions, $H_2(l)$, represents a square, and that in three dimensions, $H_3(l)$, represents a cube and so on. The unit data cube has all sides of length $l=1$, and is denoted as $H_d(l)$.

V. DECISION TREE FORMATION

The decision tree formation is completed by geographical value of data which is stored in a different item set. In order to arrive at the decision, based on the tree formation, each parent node is added with the weight; say 1 for deviation in customer related data and 2 is for deviation in transaction related data.

Further, the initial weight is incremented continuously when it travels (satisfies) through the subsequent parent node. Similarly, it maintains the weight in a same position without increment if it fails to travel through the other branches. However, the failed data is not terminated immediately rather it is allowed to travel throughout the decision tree to identify the deviation in other item sets. As the algorithm is designed to allow the travel throughout the decision tree, the combination of deviations shall be easily identified. Each particle inside the same group is classified with equal weight, as the projection of lesser value item shall rise into higher degree in money laundering and it may not be noticed if this particle value differs each other. The following plotted graph shows the several decision tree branches and their respective characteristics.

This research work is designed with the suitable algorithm which is able to identify deviations on various stages as shown above. The above pictorial representation shows the domination of samples which deviate from the samples and also the reason for increase in weight. The sample which is captured other than its parent tree normally holds the higher weight. The following decision tree shows the classification arrival based on the weight of the deviation and its respective decision rules. The algorithm segregates all these data on decision tree pattern to compare under data cube outlier analysis.

The following decision tree formation is constructed upon the decision tree rule for each partitions. Wherever the rule is matching with the similar data, the weight would get increase by means of deviation in multiple combinations and on the other hand if it is not matching with any value, the corresponding weight also in the same position but continue the traverse to find other combinations if any.

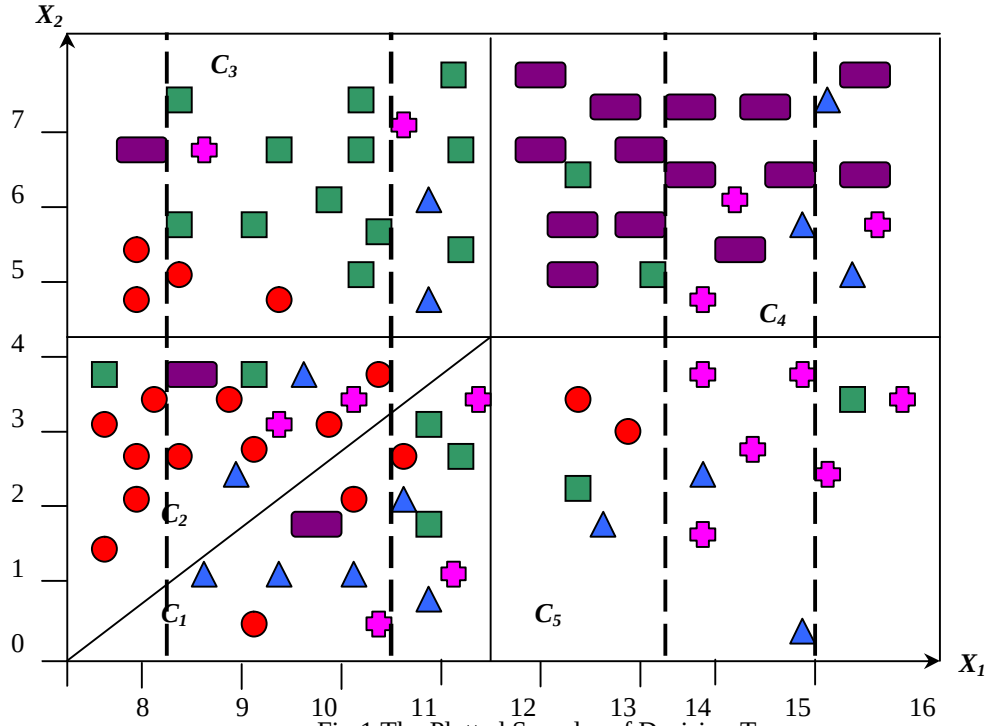


Fig 1 The Plotted Samples of Decision Tree

The below decision tree has been constructed with the help of the algorithm designed for the same which is listed below. This algorithm is dealt with each parent and child nodes of the balanced decision tree. Whenever the algorithm moves towards the construction of leaf node, a weight of the corresponding leaf also calculated and the relevant tuples are marked up with the desired flag which shall help in identifying the deviation in outlier analysis. The construction of leaf node is continuous process until it analysed all the partitions.

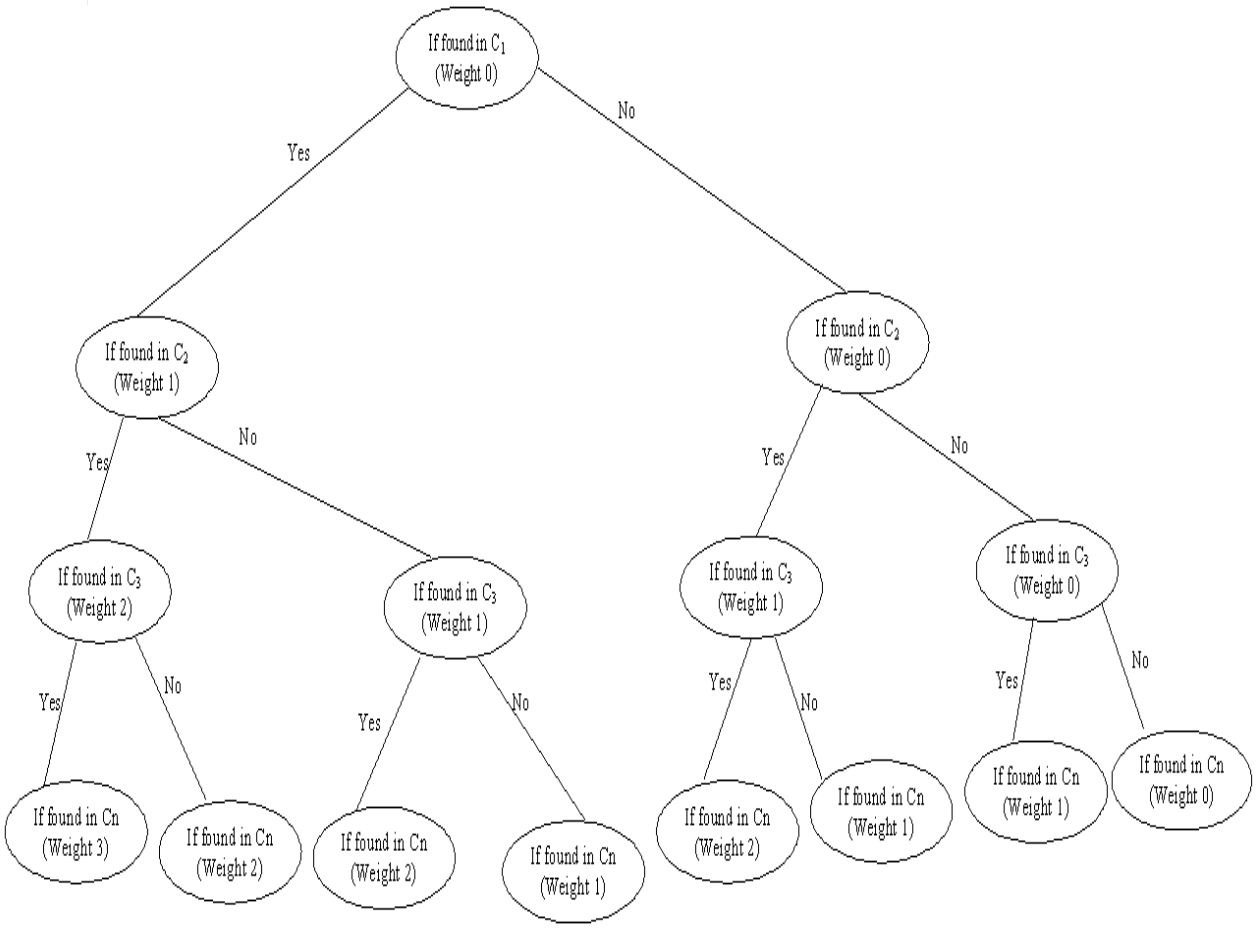


Fig 2. Formation of Rule Based Decision Tree

VI. ALGORITHM ON DECISION TREE CONSTRUCTION

n*← *[Cn] *// partition size*
n_i←***[{x_i|x_i ϵ Cn*** *, Y_i = C_i}]* ***// Size of the Attribute and class***
weight← ***(0)***
flag← ***[N]***
foreach i into n
if n <= i or i>n then ***// Termination Clause***
*Cn** ← ***arg i*** *// Leaf node creation*
return
foreach (attribute $X_{i)}$ do
if (attribute X_i = true) then
weight=weight + 1 ***// Weight adjustment***
flag← ***[Y]*** ***// Mark it as Money Laundering***
end if
end foreach;
next attribute ***// Repeat the leaf node construction***
end if;
end foreach;

VII. OUTLIER ANALYSIS – DATA CUBE BASED METHOD

As we mentioned in the introduction, the data cube shall be constructed with *'n'* number of similar cubes based on the number of attributes and partitions. Each dimension shall be enabled to express the deviation. This research work is constructed with the above ideology which shall be migrated to higher/lower dimension based on the attributes and partitions. The construction of decision tree and marking of relevant shapes/flags deploys the data cube which shall easily identify the exceptional behavior.

The Figure 3 shows the pictorial representation reflects the different surfaces of the cube where the user can identify the samples which are different from others (i.e) the account number or customer which/who may cause the money laundering. The cubical surface shall also be possible to distinguish based on the cluster and the nature of deviation/weight. The following is the algorithm which shall deploy the data cube model for the outlier analysis.

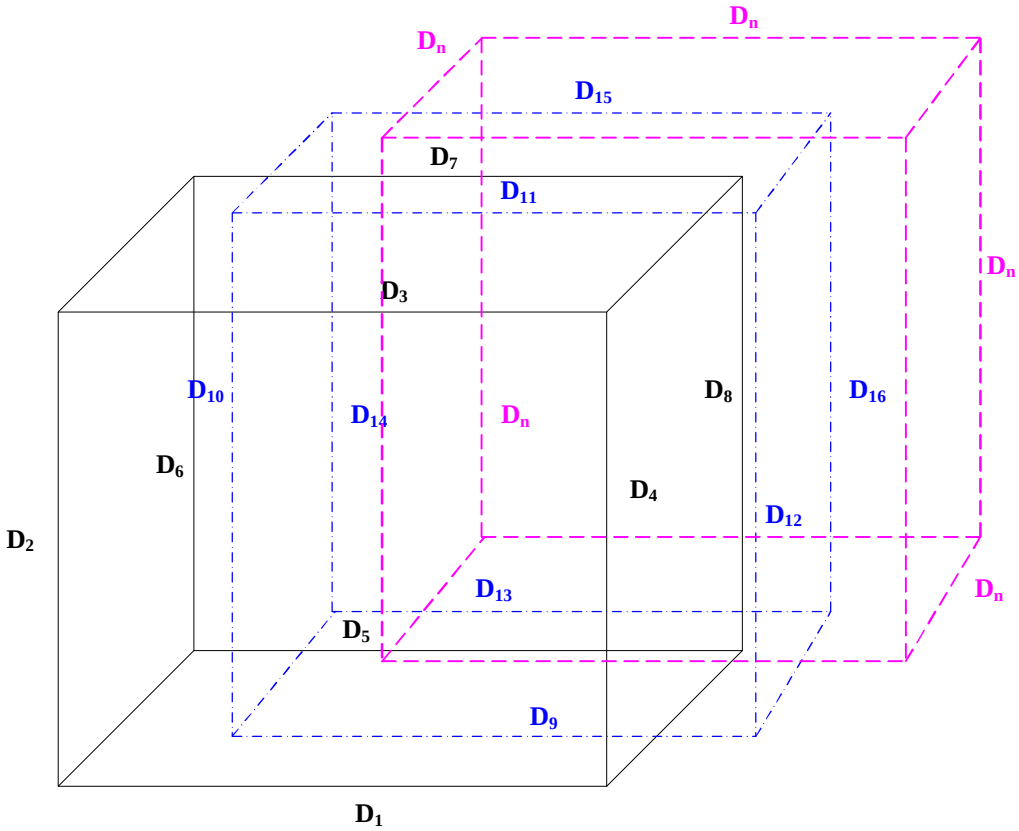


Fig 3 Data Cube Outlier Analysis

VIII. ALGORITHM FOR DATA CUBE OUTLIER ANALYSIS

```
n← [Cn]  // partition size
n_i←[{x_i|x_i ϵ Cn , Y_i = C_i}]  // Size of the Attribute and class
weight← (0)
flag← [N]
foreach i into n
if n <= i  or i>n then  // Termination Clause
   read Cn*
   if weight!=0 and flag!='N' then // Significance
   sort (weight);
   Cube* ← (arg i,weight)  // Surface of the cube segregation and identification
   Cube=&cube + 1; //Moving to next surface
   end if;
```

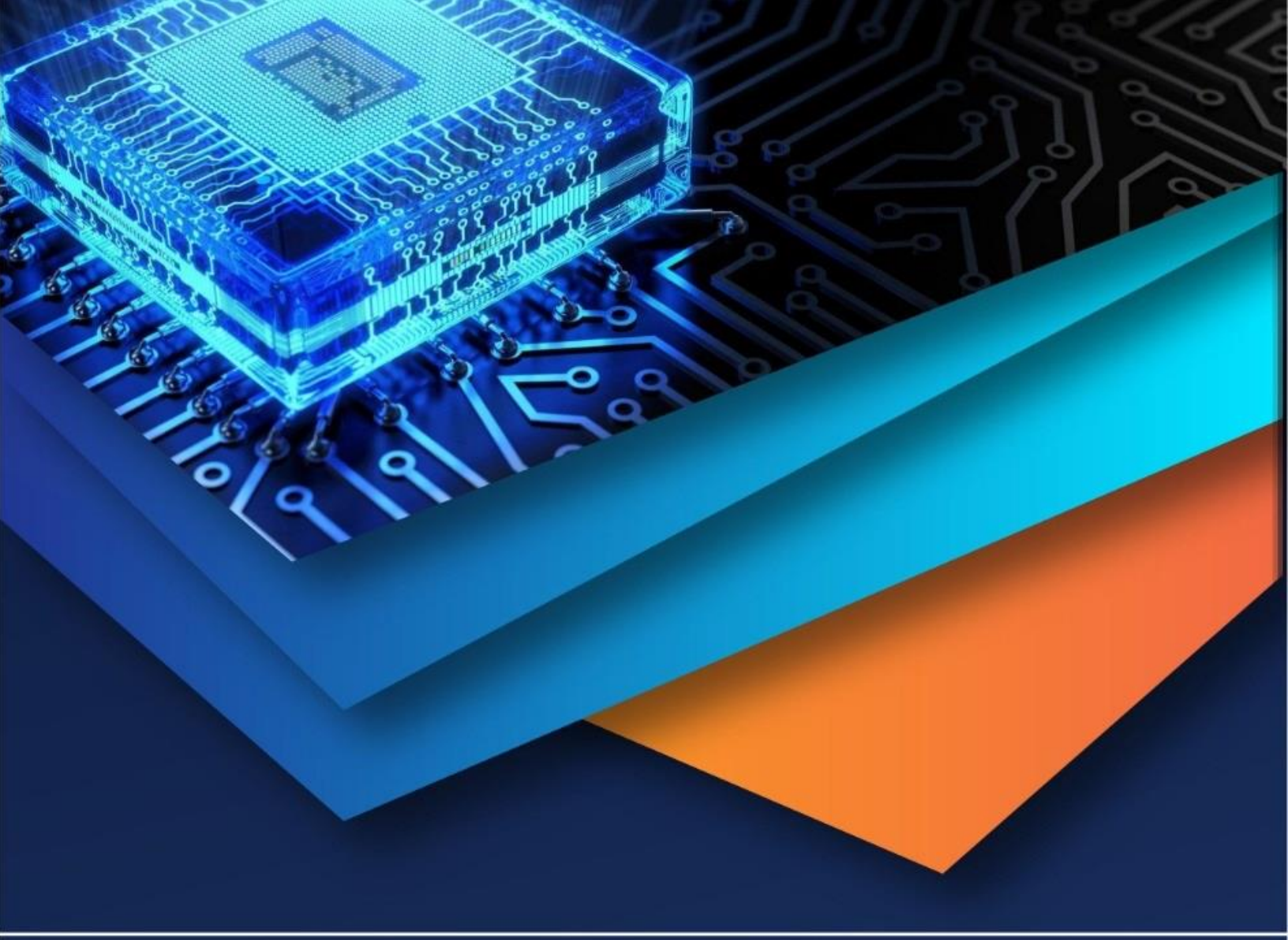
return
next attribute ***// Repeat for next cubical surface***
end if;
end foreach;

IX. CONCLUSION

The above algorithm is matching all the attributes based on the weight and the flag value to assign relevant surface under the data cube. The weightage of the risk value is the significant and constructed under the data cube algorithm. Further based on this, the priority is fixed as high or medium or low level during the analysis of data cube. The user shall view the data cube surface based on the weight or cluster or key identifiers. The data cube shall also to be migrated based on the above values for the user defined sets.

REFERENCES

[1] R. Agrawal and R.Srikant, "Fast Algorithms for Mining Association Rules," Proc. 1994 Int'lconf. Very Large Data Bases, pp. 487-499, Santiago,Chile,Sept. 2010.
[2] Haruka Fuse, Haruka Fukamachi, Mitsuko Inoue and Takeshi Igarshi, "Identification and Functional Analysis of the Gene Cluster",Gene Volume 515,Issue 2, Pages 291-297,25th February 2013, Elsevier Publications
[3] D.W. Cheung, J. Han, V. Ng,A. Fu, and Y. Fu, "A Fast Distributed Algorithm for Mining Association Rules," Proc. 1996 Int'l Conf. Parallel and Distributed Information Systems, PP. 1996 Int'l Conf. Data Enf., PP. 106-114, New Orleans, Feb. 2010.doi 10.1109/PDIS.1996.568665
[4] L. Li, C.R. Weinberg, T.A. Darden, L.G. Pedersen, "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method", Bioinformatics 17 (12) (2001) 1131–1142.doi.10.1007/978-3-642-13089-2_49
[5] J. Khan, J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C.R. Antonescu, C. Peterson, P.S. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks", Nat. Med. 7 (6) (2001) 673–679.
[6] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Bostein, "Cluster analysis and display of genome-wide expression patterns", Proceedings of the National Academy of Science USA 95 (1998) 14,863–14,868.
[7] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. GaasenBeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri,C.D. Blomfield, E.S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene-expression monitoring", Science 286 (1999) 531–537.doi.10.1126/science.286.5439.531
[8] E. Frank, I.H. Witten, "Generating accurate rule sets without global optimization", in: , Machine Learning: Proceedings of the 15th International Conference, Morgan Kaufmann Publishers, Los Altos,CA, 1998
[9] Y. Fu and J. Han, V. Ng, A. Fu, and Y. Fu, "A Fast Distributed Algorithm for Mining Association Rules," Proc. 1996 Int'l Conf. Parallel and Distributed Information Systems, PP. 31-44, Miami Beach, Fla.,Dec. 2001.
[10] D.W. Cheung, J. Han,V. Ng, and C.Y. Wong, "Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique," Proc. 1996 Int'l Conf, Data Engg., PP. 106-114, New Orleans, Feb. 2009.doi.10.1109/ICDE.1996.492094
[11] D.W. Cheung, J. Han, V. Ng,A. Fu, and Y. Fu, "A Fast Distributed Algorithm for Mining Association Rules," Proc. 1996 Int'l Conf. Parallel and Distributed Information Systems, PP. 1996 Int'l Conf. Data Engg., PP. 106-114, New Orleans, Feb. 2010.doi.10.1109/ PDIS.1996.568665
[12] M.S. Chen, J. Han, and P.S. Yu, "Data Mining: An overview from a Database Perspective," IEEETrans. Knowledge and Data Engg., Vol.8, PP.866-883,1996
[13] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," Proc. 1993 ACM SIGMOD Int'l Conf. Management of Data, pp. 207-216, Wahington, D.C., May 1993.doi.10.1145/170036.170072



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)