

Apprehension and Analysis of Pollution Under Control Certificate using ID3 algorithm greedy method

Dr. Neeru Bhardwaj ,Asso.Prof,
Rayat Bahra University.edu.in
neeru.cse@rayatbahrauniversity.edu.in

Er Harpreet Singh,Asst..Prof.
Rayat Bahra University.edu.in
harpreetsingh.11995@rayatbahrauniversity.edu.in

ABSTRACT: Inductive learning was found to depend on acceptance. The basis of induction research is the decision tree method. This method uses inductive methods to correctly classify objects based on the quality they provide. These calculations are very useful for characterizing articles and are often used in the main frame. When any new quality is added, the ID3 decision tree learning calculation is used to see if there is any progress in the selection principle currently granted to Pucc (Pollution Treatment Certificate). Three studies were conducted on the Pucc issuance, each of which added a new feature to the data set, thereby providing decision rules.

KEYWORDS: - inductive, classification, Pucc, ID3, U, N, Y,N,Ent, decision learning algorithm

ID3, or Iterative Dichotomize 3, is a decision tree learning computation that is used in the iterative inductive process to determine the order of the elements. The top down approach is used in calculation. The root node is the top most node in the tree. The other nodes are called leaf nodes. As a result, it is a cross from the root node to the leaf node. Each node requires a test of the criteria that determine the rank of the leaf node. Most of the time, these decision trees are used for dynamic reasoning [7].

Decision tree learning is a strategy for determining why having the right amount of capacity is worth it. Decision trees are used to indicate learned abilities. For inductive reasoning, decision tree learning is one of the most common and widely used strategies and is practical in practice[2][3].

In light of the three reasons, decision tree learning calculations were used:

1. Decision tree is a great way to get information from rare cases.
2. The calculations in these techniques are effective and in accordance with the observed conditions.
3. Finally, humans can easily understand the critical tree that was built. [1] [2]

Defines decision rules using greedy bottom-up search. This requires the use of statistics. ID3 is primarily concerned with the two concepts of ENTROPY and INFO GAIN.

The unpredictability of each factor is calculated using entropy. The entropy formula is as follows:

$$n^{\text{th}} \text{-class Ent : } E(S) = \sum - (p_i * \log_2 p_i)$$

$$2^{\text{nd}} \text{-class Ent : } E(S) = - (p_1 * \log_2 p_1 + p_2 * \log_2 p_2)$$

P(positive) are the positive examples in S

P(negative) are the negative examples in S

S is the example preparing set having positive and negative qualities.

Entropy can also be defined as an element of contamination in the employment preparation set [1].

The decisive tree that emerges after determining the entropy is very large.

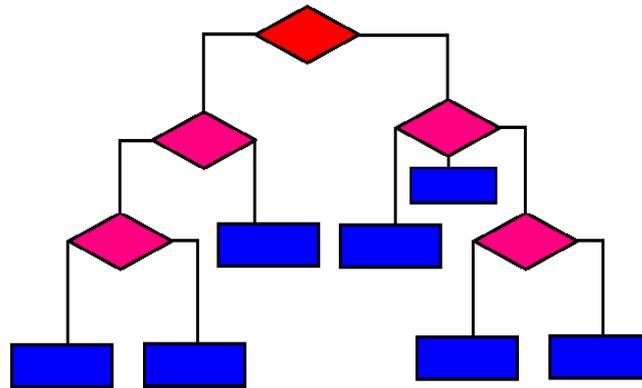


Figure 1: Decision tree after entropy is calculated.

Obtaining information

ID3 uses information gain to separate the standards of nodes. The root node will be an additional mass. The IG formula is as follows:

$$\text{Gain}(S,A) = E(\text{before}) - G(\text{after_splitting})$$

The next tree will grow when IG decreases the depth of the tree.

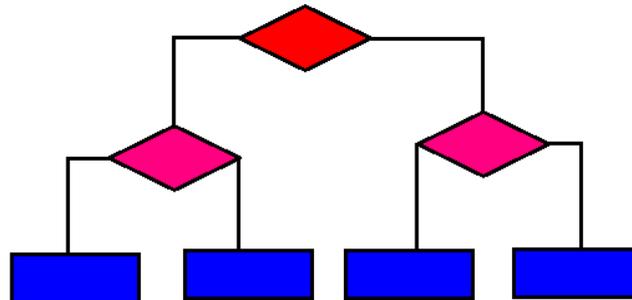


Figure 2: Decision tree with reduced depth after IG is calculated.

This technique is repeated indefinitely until the desired classification is achieved.

Classification rules.

If the attribute's entropy is 0, the nodes are homologous and do not need to be further classified.

Assuming the entropy of the attribute is 1, the nodes are heterogeneous and there is a strong reason to further characterize them.

LITERATURE REVIEW

J.R. Quinlan was the first to work on ID3 in 1986. It uses the decision tree which has been used in many systems, which they call ID3 [6]. In ID3, data mining processes are used as the primary classification algorithm. Milija Suknovic (et al.) Discovered some ID3 algo reusable components. They merged reusable elements to duplicate the actual method. They also discovered that their changes would create new induction algorithms [13].

On a side note, Victor H.Goreia (et.al) employed ID3 network security to detect online attacks. It makes decision rules that are easy to categorize and understand the source of the attack [8]. Another researcher, Sonica Tiwari, used ID3 generated using a deciduous tree on a horizontal section to discover network anomalies. [12].

The ID3 algorithm is widely used in data mining methods because it is the most important permutation algorithm. ID3 is mainly used for information mining in the field of healthcare. Hu Ruijuan used ID3 method to calculate breast cancer growth information to predict the relationship between repetitive features and different breast cancer features [9].

L. Satish Kumar and A.Padmapriya have also completed the prediction of common illnesses on cellphones and television using ID3. They offer an approach that helps people learn about diseases in order to reduce the rate of transmission and the number of people affected by the diseases [15].

Victor H.Goreia did an interesting job of using ID3 in the realm of organizational security. This creates selection criteria that are easy to describe and understand as the basis for the attack [8]. Sonika Tiwari was another expert who used the optimal ID3 network to identify anomalies with a decision-based distribution tree [12].

Mary Slocum used ID3 data mining in her medical studies. She makes changes and transforms large volumes of data into information that can be used to make decisions [10].

Some writers have developed and evaluated ID3 in various databases. For example, Anand Bahety implemented a method in the "PlayTennis" database. Did he identify the ideal weather conditions for tennis? [9]. The survey was conducted using the basic "loan application" dataset by T.Y HSU 662096093. He decided if an individual would be eligible for a mortgage. Kumar Ashok (et al.) Followed. Using the ID3 method, he categorized the consensus. Another study used the same method to determine if a person was tanned. [11]

This article uses the ID3 example to show the requirements for issuing a PUCC to a car. The ID3 algorithm is applied to the three databases to obtain the decision criteria. The question is whether the car will receive a PUCC?

OBJECTIVE OF THE RESEARCH WORK

The main research objectives are:

- i. Make a decision tree for proper classification.
- ii. Another option is to make rules for problem solving.

RESEARCH METHODOLOGY

The appropriate approach used to arrive at the conclusion of the research effort.

All information about the problem is contained in the dataset that was used (attributes, their values, etc.). For testing purposes, three datasets were developed. A new attribute has been added to each dataset. The information gain is estimated after the dataset has been produced and evaluated for the entropy calculation. [The procedure is performed for each data set separately.](#)

The last phase is to create "decision trees" and "decision rules". The three studies are then compared to see what has changed in the current situation

RESULTS AND DISCUSSION

The current state of PUCC issuance is a one-time manual validation for measuring pollution using physical machinery, and the data base contains only the values of pollutant produced by the vehicle. So far, the only way to get a PUCC is to do a manual check.

Present Rule is:

```
IF manual check up =approved
THEN
Issue= Y
ELSE
```

Issue=N

Therefore, the purpose of this article is to amend this rule by adding some more points that determine whether PUCC is issued or not. Here in this inspection, it will be determined whether the vehicle has some marks in the manual inspection to determine whether the vehicle is PUCC approved.

The final features of the dataset are determined after pruning the various properties associated with the problem.

1. The domain values of the fuels attribute are diesel, CNG (compressed natural gas) and petrol.
2. The domain values of the category attribute are two wheeler maker, Three ships and a four-wheeler
3. Because any vehicle's minimum distance is set to 5000 kilometers, two node characteristics will be produced to indicate whether or not they have travelled. As a result, Fatigue or Insured will be the domain of the vehicle for attribute Kilometer.
4. The vehicle will be serviced within three months of receiving the PUCC, therefore the domain of the vehicles attribute Service will be Yes or No.
5. When it comes to any car that has been in production for more than 15 years, according to the authorities, is unfit for transportation. This could be a key factor in the PUCC's decision in this case the attribute Year has domain value will be true or false
6. For any sub features, run manual checks the attributes are Approved and Not approved
7. The target feature has the node features as an issue is Yes and No

Now the three studies are discussed and their Decision trees and decision rules are generated below.

STUDY 1

Table I

Name	Fuel	Category	Kilometer	MC	Issue
Anshu	Pertrol	Two	Covered	Pass	Yes
Chirag	Pertrol	Two	Not Covered	Pass	Yes
Ruzal	Pertrol	Four	Covered	Fail	No
Tanishq	Pertrol	Four	Covered	Pass	Yes
Ravi	Diesel	Four	Covered	Fail	No
Ruhi	Pertrol	Four	Not Covered	Fail	No
Purnima	Diesel	Three	Covered	Fail	No
Somel	Diesel	Three	Covered	Pass	Yes
Raman	CNG	Three	Not Covered	Pass	Yes

Here the kilometer is the feature which is added to the present scenario and the study decision tree and decision is as shown in Figure 3

On the basis of rules were constructed.

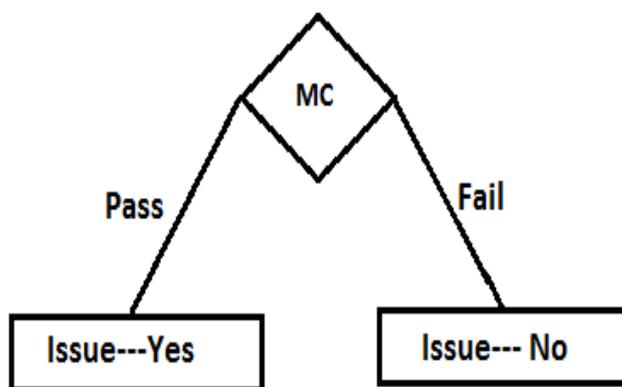
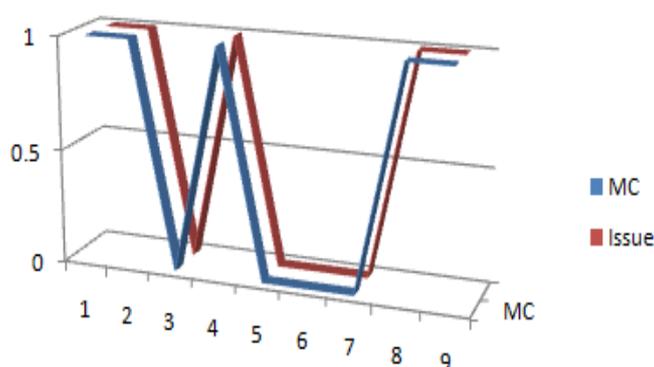


Figure 3: Decision tree for study 1

The rules created after adding the km attribute are

```

If (MC=="P"){
    Issue=="Y";
} elseif(MC=="F") {
    Issue=="N";
}
    
```



Graph: 1

STUDY 2

Table II

Name	Fuel	Category	Kilometer	Service	MC	Issue
Anshu	Pertrol	Two	Covered	Yes	Pass	Yes
Chirag	Pertrol	Two	Not Covered	No	Pass	Yes
Ruzal	Pertrol	Four	Covered	No	Fail	No
Tanishq	Pertrol	Four	Covered	Yes	Pass	Yes
Ravi	Diesel	Four	Covered	No	Fail	No
Ruhi	Pertrol	Four	Not Covered	Yes	Fail	No
Purnima	Diesel	Three	Covered	No	Fail	No
Somel	Diesel	Three	Covered	No	Pass	Yes
Raman	CNG	Three	Not Covered	No	Pass	Yes

Here the service is the attribute which is added to the present scenario and the study decision tree and decision is as shown in Figure 3

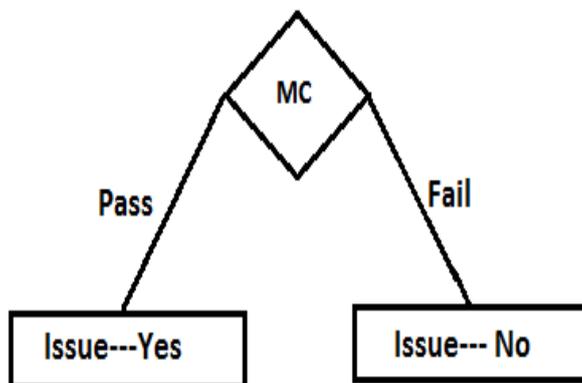
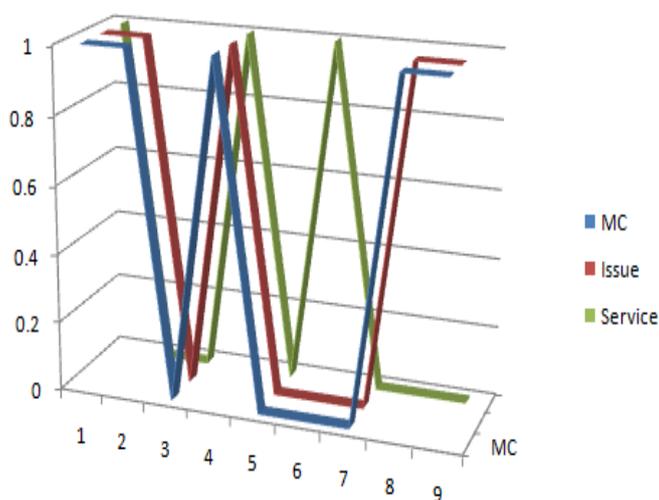


Figure 4: Decision tree for study 2

The rules created after adding the kilometer attribute are

```

If (MC=="P")
{
    Issue=="Y";
}
elseif(MC=="F")
{
    Issue=="N";
}
    
```



Graph: 2

STUDY 3

Table III

Name	Fuel	Category	Kilometer	Service	Year	MC	Issue
Anshu	Pertrol	Two	Covered	Yes	Invalid	Pass	Yes
Chirag	Pertrol	Two	Not Covered	No	Valid	Pass	Yes
Ruzal	Pertrol	Four	Covered	No	Valid	Fail	No
Tanishq	Pertrol	Four	Covered	Yes	Invalid	Pass	Yes
Ravi	Diesel	Four	Covered	No	Valid	Fail	No
Ruhi	Pertrol	Four	Not Covered	Yes	Invalid	Fail	No
Purnima	Diesel	Three	Covered	No	Valid	Fail	No
Somel	Diesel	Three	Covered	No	Valid	Pass	Yes
Raman	CNG	Three	Not Covered	No	Valid	Pass	Yes

The year is a new component that has been added to the existing scenario.

On the basis of the study, decision tree and decision rules have been developed.

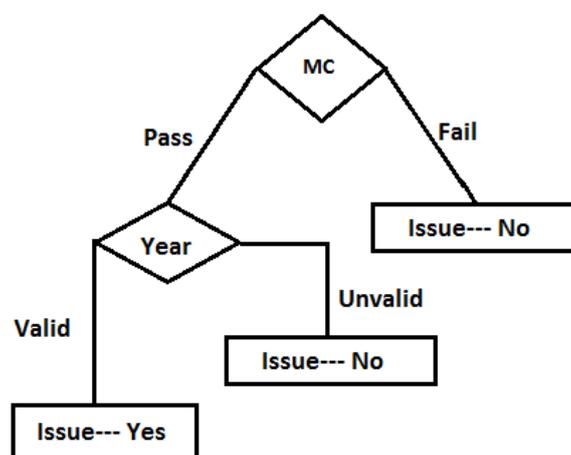
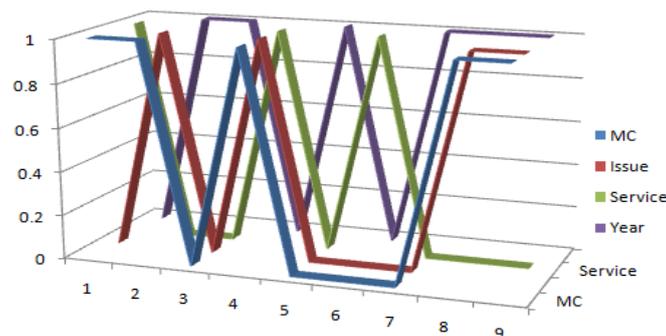


Figure 5: Decision tree for study 3

The rules created after adding the kilometer feature are

```

If (MC=="P")
{
  If (Year=="V")
  {
    Issue="Y";
  }
  elseif (Year=="U")
  {
    Issue="N";
  }
  elseif (MC=="F")
  {
    Issue="N";
  }
}
  
```



Graph: 3

VI. CONCLUSION

We deduce from this work and its application that datasets with various values are needed for the ID3 decision tree learning algorithm to work successfully [7]. According to the findings of the study, the year is a characteristic that is also taken into account during a manual inspection on automobiles before granting a PUC.

As a result, the revised decision-making criteria are as follows:

```

If (MC=="Pass")
{
    If (Year=="Valid")
    {
        Issue="Y";
    }
    elseif (Year=="U")
    {
        Issue="N";
    }
    elseif (MC=="Fail")
    {
        Issue="N";
    }
}

```

So today, whenever a person forgets PUC, his car must pass the vehicle's age test as well as a physical test. He will only be eligible for PUC if his car is under the age of 15 and has passed a manual inspection. This will reduce road pollution and traffic to some extent..

FUTURE SCOPE

Attempts will be made at a higher level to identify hidden contaminants that have not yet been assessed in the current system, according to the current study. Researchers can work to reduce pollution caused by sulfides and nitrides released from vehicles by developing new equipment or software.

REFERENCES

- [1] Paul E. Utgoff and Carla E. Brodley, (1990). "An Incremental Method for Finding Multivariate Splits for Decision Trees", *Machine Learning: Proceedings of the Seventh International Conference*, (pp.58). Palo Alto, CA: Morgan Kaufmann
- [2] Wei Peng, Juhua Chen and Haiping Zhou, "An Implementation of ID3 - Decision Tree Learning Algorithm". University of New South Wales, School of Computer Science & Engineering, Sydney, NSW 2032, Australia
- [3] Tom M. Mitchell, (1997). "*Machine Learning*", Singapore, McGraw-Hill.
- [4] <https://towardsdatascience.com/machine-learning-101-id3-decision-trees-and-entropy-calculation-1-a1d66ee9f728>
- [5] <http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.html>

- [6] *Quinlan, J.R. 1986. "Induction of Decision trees Machine Learning"*
- [7] Anand Bahety, "Extension and Evaluation of ID3 – Decision Tree Algorithm".
- [8] Victor H. Garcia, Raul Monroy and Maricela Quintana, "Web Attack Detection Using ID3". <http://homepage.cem.itesm.mx/raulm/pub/ifip06.pdf>.
- [9] Ruijuan Hu,(2011)." Medical Data Mining Based on Decision Tree Algorithm", Computer and Information Science ,Vol. 4, No. 5; September 2011.
- [10] Mary Slocum,"Decision making using ID3 algorithm". Rivier academic journal, volume 8, number 2, fall 2012
- [11] *Kumar Ashok, Taneja H C, Chitkara Ashok K and Kumar Vikas, "Classification of Census Using Information Theoretic Measure Based ID3 Algorithm" . Int. Journal of Math. Analysis, Vol. 6, 2012, no. 51, 2511 – 2518.*
- [12] Sonika Tiwari and Prof. Roopali Soni, "Horizontal partitioning ID3 algorithm A new approach of detecting network anomalies using decision tree". International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 7, September – 2012.
- [13] Milija Suknovic, Milos Jovanovic, Boris Delibasic , Milan Vukicevic, Dragana Becejski-Vujaklija, Zoran Obradovic (2011). "Reusable components in decision tree induction Algorithms".
- [14] Herbert Schildt, *Java 2: "The Complete Reference, 5th edition"* , McGraw-Hill/Osborne
- [15] L. Sathish Kumar and A. Padmapriya, "Prediction for Common Disease using ID3 Algorithm in Mobile Phone and Television". *International Journal of Computer Applications (0975 – 8887) Volume 50 – No.4, July 2012.*
- [16] <http://www.wikipedia.com/>