ID3 Algorithm to Identify Customer Loyalty Factor at Semarang Ceramics Company

Isadora Nugroho
Master of Information Systems,
Faculty of Information
Technology, Satya Wacana
Christian University, Indonesia

Danny Manongga
Information System Department
Faculty of Information
Technology, Satya Wacana
Christian University, Indonesia

Wiranto H. Utomo
Information System Department
Faculty of Information
Technology, Satya Wacana
Christian University, Indonesia

ABSTRACT

The aim of this research is to identify the customer loyalty at a Semarang ceramics company. The research process uses ID 3 algorithm and 5 SERVQUAL attributes, *reliability*, *assurance*, *tangible*, *empathy*, and *responsiveness.* Questioner data is the main data, which is analyzed by WEKA 3.7.7 software. The result of this research is Responsiveness attribute and its indicator quick service to the customer is the main factor which is influence the customer loyalty.

Keywords
Decision Tree, Algorithm ID3, Service Quality, Customer Loyalty

1. INTRODUCTION
Recently, ceramics import from other country enters Indonesia and they give the best offer to the customer. Mirrored to this situation, local company should prepare and plan some strategic in providing the best service to the customer. With good strategic, local company can compete to foreign ceramics company especially to reach the customer loyalty. This research observes the problem in Semarang Ceramics Company, Indonesia. There are some factors to reach customer loyalty, first, creating a good relationship to the customer, give the best service and maintain the customer satisfaction. To implement these three factors is not as easy as turning a hand. Company should exert every effort to reach the customer satisfaction and profit company so the company can survive in the middle of competition.

Company should prepare and plan the best strategy to face the competition and to reach the potential customer. There are 5 dimension which influence the customer loyalty, reliability, assurance, tangible, empathy, and responsiveness. In measuring customer satisfaction and customer loyalty, analyzed and data observed should be held to find the factors of the problem through quisioner method. Some journals use ID3 Algorithm as decision tree in every problem. ID3 concept implementation could help the company to identify potential customer. From the tools, the company can understand the customer behavior as a recommendation to marketing implementation strategy which can reach the best benefit to Semarang Ceramics Company.

2. LITERATURE REVIEW
There are some previous researches related to the theory. The first title is ID3 Classifier for Pupils' Status Prediction [8]. The research explained that decision tree is used as a basic to predict required student status to further studies. Data mining is the source data of this research. The result of this research is the student classification which has high academic output.

Second research evaluated the patient disease diagnose, Performance Evaluation of Decision Tree Classifiers on Medical Datasets [1]. This research is a broad material, so the researcher decide to use decision tree to diagnose disease such as Brest cancer diagnose with ultrasonic picture, ovaries cancer and heart beat diagnose. The research compares analysis accuracy and time complexity by using ID3, C4.5 and CART. These tools is able to read and translate to medical diagnose. The result of this research is decision tree works well in the medical classification data.

The last research is Case Study on High Dimensional Data Analysis Using Decision Tree Model [12]. This research is to predict the disease possibility in a region. This research, which identificate the parameter significant in prediction process, uses Decision tree. There are 3 models which is created trough ID3 Algorithm, unsupervised model to identification low level family characteristics and supervised model to high level family. Grouping model is enabling to understand the disease history group and decision tree is used to solve problem. The result is risk factor like climate, rainfall, deadly disease spread, water, temperature and environment is the highest factor which influences the low level family life and hereditary health history is for high level one.

In calculating and creating decision tree, this research uses ID3 Algorithm. Customer questionnaire result, with 5 dimensions in SERVQUAL, is the research data. In calculation process, the system is automatically calculating in ID3 Algorithm. The result of this research is loyalty customer format, loyal and non-loyal customer to the company.

3. CUSTOMER ANALYSIS
This research focuses on customer analysis especially customer behavior. The aim is to identify the loyalty customer factors and its relation to customer service quality of the company. These factors are divided to some variable and attribute which is the most influence to the company. From the variable data, it can conclude basic factor of customer loyalty to the company.

3.1 Customer Loyalty
Customer loyalty is repurchasing of a product or a service because of the interest of the product or the service. Customer loyalty is a continue manifest of customer satisfaction in using a product or service. There are some loyal customers like below [6]:

1. Regularly purchase in the company. A loyal customer repurchases repeatedly in the same company in a period.

2. Purchase out of common product/service. Loyal customer doesn't only purchase a product, but also purchase another product or service in the same company.

3. Give a recommendation to others. Loyal customer will recommend to others about their positive experience in purchasing a product or a service in a company. They persuade other not to purchase in the other company.

4. Show uninterested to try another product. They trust the product or a service, that they use is the best product and different to other.

3.2 Service Quality

A group of attribute is needed to analyze service quality. Attribute is behavior or specific characteristics which represent and point of the service [4]. Service dimension is the main thought in deciding the characteristics of the service. There are 5 dimensions to analyze the service:

1. Tangible
Attribute and dimension, which include physical facilities, to fulfill the customer satisfaction, like: building, equipment.

2. Reliability
Attribute and dimension, which include capability in servicing to the customer like promised.

3. Responsiveness
Willingness to help customer and responsive to what customer need. Capability to give proper information and always give the best service as soon as possible.

4. Assurance
Make a safety feeling and comfortable to the customer. These conditions influence to the customer because of its credibility, politeness, hospitality and employee knowledge in servicing customer.

5. Emphaty
Give more attention to the customer. Simplicity of making a relation, good communication, personal attention and understand the customer needs.

3.3 Relation between Service Quality and Customer Loyalty

Service quality and customer loyalty has a tight correlation. If the company can fill customer's expectation, need and wish, the value of the service quality is good. They will be a loyal customer if their need, wish and expectation about the product are filled. For sure, they will come back and repurchase a product or service in the same company and don't have any interest to purchase another product from competitor.

Having loyal customer is the end of the business. To reach the aim, the company should improve their service quality with focus on 5 dimensions, *tangible, reliability, responsiveness, assurance,* and *empathy*. Through the efforts, the company can fill customer need, hope and expectation. It can attract loyal customer to come to purchase in the company, like shown in Figure 1.
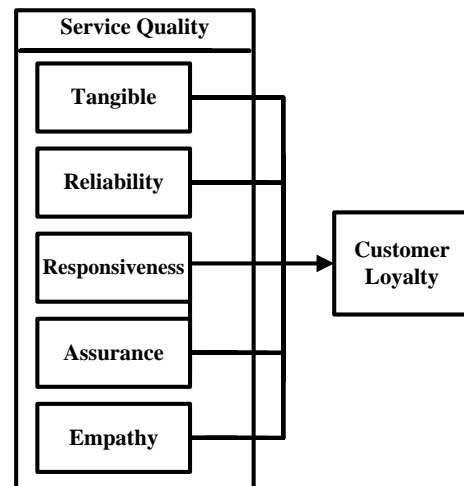

Figure 1: Correlation between Service quality and Customer loyalty

4. ID3 ALGORITHM

One of induction decision tree algorithms is ID3 (Iterative Dichotomies 3). ID3 built by J. Ross Quinlan [1]. ID3 Algorithm could be implemented to Recursive function (function which can call it). ID3 is basic decision tree learning algorithm. Algorithm searches greedily in all decision tree possibilities.

ID3 characteristics in bulding decision tree are from root to leaf (top-down) and training data recursively. It is parted to a small categories in built the tree (divide and conquer). The first attribute should be plant on the root, and then it is evaluated by statistics information gain measurement. Data group consist of decision variable and outcome variable. Both variables should have categorical value. ID3 need the value as a label and symbol which doesn't have relation one another. Below is the ID3 Algorithm:

1. Tree is started as a single root, which represent all data.

2. After node root, the data will measured by information gain to select which attribute can be divided attribute.

3. A branch of the tree is made from the divided attribute and the data will distribute to other branch.

4. Algorithm will use the same process (recursive) to make secession tree. When the attribute is the dividing node or branch, the attribute will be kept out from information gain value.

5. Recursive dividing process will stop if one of the conditions is not fulfilled:

- All data from branch include in the same class.

- All attribute is already used, but the rest is in the different class. In this case, we should take the data which represent the most class to be class label in the node leaf.

- There is no data on the new branch. In this case, we should take from the previous node and the data is

taken which represent the most class to be class label.

4.1 Entropy

A classified object on the tree should be tested entropy value. Entropy is a measurement of the information theory to get the characteristics of impurity and homogeneity from the group data. From entropy value, we can count the information gain value (IG) from each attribute [8].

$$\text{Entropy}(S) = -p_a \log_2 p_a - p_b \log_2$$

S : Sample Room (data) to *train*.
P_a : Positive Solution amount (supported) of the sample data in some criteria.
P_b : Negative Solution (unsupported) of the sample data in some criteria.

From the entropy formula show that Entropy (S) is a needed bit to be extracted in a class (a or b) from the random data in an S sample room. It can conclude that the smaller entropy value, the better extraction in a class. The code, which gives optimal information, is $-\log_2 p$ bits. It is to inform probability p. Bit amount to extract S in to a class is $-p_a \log_2 p_a - p_b \log_2$.

4.2 Information Gain

After getting entropy value in a group data, we can measure the attribute affectivity to classify the data. It is information gain. Information gain of an attribute is like bellow [8]:

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

A : attribute
v : a possibility value for an attribute
Values(A) : a possibility assemblage for an attribute
|Sv| : Sample amount for v value
|S| : Total sample data
Entropy(Sv) : *entropy* for sample v value

5. DATA PREPARATION AND ARCHITECTURE MODEL

The method of collecting data is questionnaire method. The questionnaires spread out to the customer, which is direct purchase and online purchase in a Semarang ceramics company. There are 128 data collected and will be analyzed. The analysis is to solve the problem which shows up in the company, especially Service quality dimension. Flowchart model as shown in Figure 2.
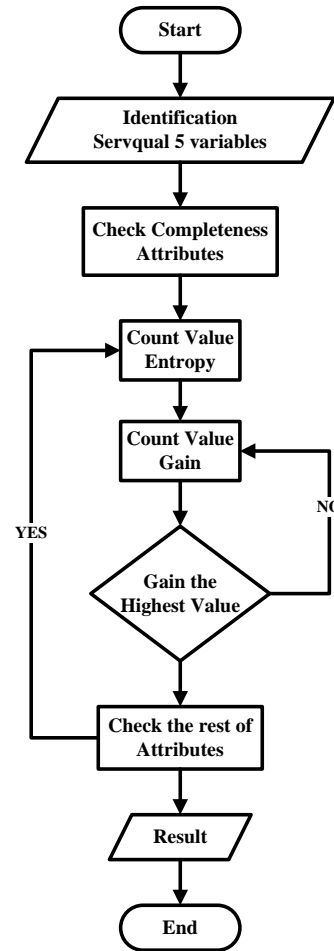

Figure 2: Architectural Model Flowchart

6. DISCUSSION

The analysis method to implement ID3 Algorithm based on the scope of company service quality especially in Semarang Ceramics Company to the customer. Qualitative approach is used to make quantitative approach. Case study of this research is the service quality implementation can be seen in Figure 3.
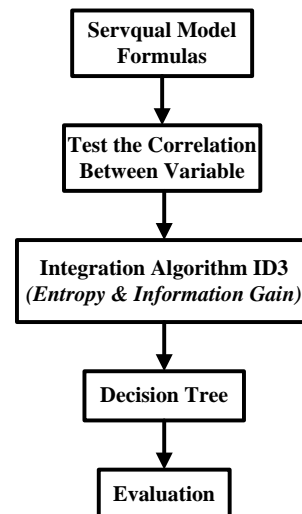

Figure 3: Servqual Method with ID3

6.1 Servqual Model Formulas

Model Servqual variable is like below:

1. Tangibles (X1) surface of service quality like physical facilities, equipment, personnel and communication.

2. Reliability (X2) Capability to show the best service like promises punctual and trustable. Service should be punctual and in time without any fault.

3. Responsiveness (X3) willingness to help the customer and give a proper service. Long service cause negative response about the company service.

4. Assurance (X4) knowledge, politeness and capability to say trust and assure to the customer and give safety feeling or guarantee to the service.

5. Empathy (X5) give attention and understanding to customer personally as shown in Table 1.

Table 1. Variable in Servqual Model

Variable	Operational	Indicators
Tangibles (X1)	Real aspect of the physical appearance.	- Product - Modern equipment - Facilities - Employee Neat
Reliability (X2)	Capability to serve on time, punctual and trustable	- Consistent - On time - Serve like promised - Care - Sincerity
Responsiveness (X3)	Willingness to help customer and give a	- Clear Information
	proper service.	- Service Speed - Service accuracy - Willingness to help - Readiness to respond
Assurance (X4)	Knowledge, politeness and employee capability to say trust and say trust to the customer. Give safety feeling or guarantee.	- Trust - Safe - Comfortable - Polite - Guarantee
Emphaty (X5)	Give a personal attention to the customer.	- Personal Attention - Proper operational time - Specifically need understanding - Work time suitability

In the research, service quality variable will represented trough alternative respond like: Good, Bad, Quick, Slow, Convince, Satisfaction, Comfortable and Less.

6.2 Correlation Test between Variable

If attribute and variable are correlated, the relation can be shown in Figure 4.

Figure 4: Correlation Test between Variable

6.3 ID3 Algorithm Integration (*Entropy & Information Gain*)

Target Attribute is customer loyalty which has loyal and disloyal value, while dimension is Reliability, Responsiveness, Assurance, Empathy, and Tangibles. There are 128 services from customer types, and the loyalty can be shown in Figure 5.
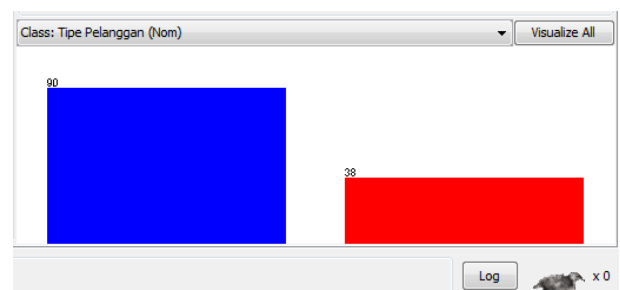

Figure 5: Target Attribute

There are 90 types of loyal customer and 30 types of disloyal customer. The visualization can be shown in Figure 6.
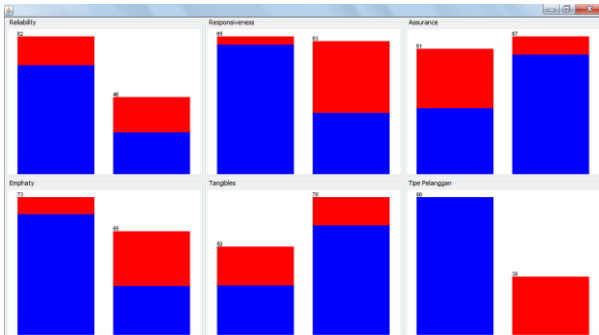
Figure 6: Servqual Attribute Dimension

6.3.1 Reliability Dimension

There are 64 well worth variable from loyal type. 33 well worth variables for disloyal type, 18 variables with a bad value for loyal type and 13 variables with a bad value for disloyal type.

6.3.2 Responsiveness Dimension

There are 50 quick value variables from loyal type, 13 quick variable from disloyal type, 15 slow value variable from loyal type and 50 slow value variables from disloyal.

6.3.3 Assurance Dimension

There are 29 low value variables from loyal type, 12 low variables from disloyal, 32 convince value variable from loyal type and 55 convince value variable from disloyal.

6.3.4 Emphaty Dimension

There are 61 satisfy value variables from loyal type, 25 satisfy value variables from disloyal, 12 less satisfy value variable from loyal type and 30 less satisfy value variable from disloyal.

6.3.5 Tangibles Dimension

There are 21 less comfortable value variables from loyal type, 22 less comfortable value variables from disloyal, 29 comfortable value variable from loyal type and 30 comfortable value variable from disloyal.

6.4 Decision Tree
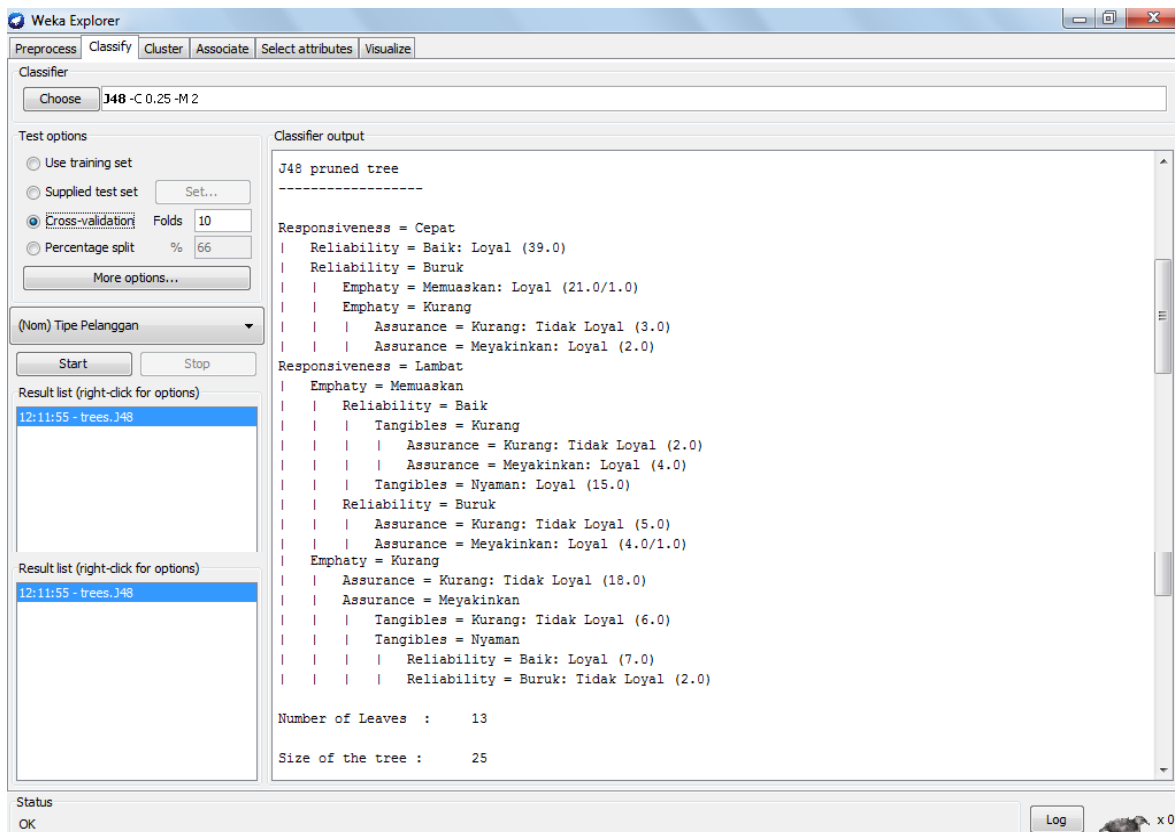
Figure 6: Tangibles Attribute

Figure 6 is a tracking decision tree problem trough ID3 Algorithm. It shows that Responsiveness is a root of the tree. The tree has some leaf, Reliability, Empathy, Assurance. Empathy, Reliability, Tangibles, Assurance. Empathy, Reliability, Assurance. Empathy, Assurance, Tangibles, Reliability.

6.5 Evaluation

From the decision tree, there is ID3 algorithm which shows the customer loyalty pattern from Service quality point of view. Some factors which cause the loyal customer can be seen from Figure 7.

```
Ressponsiveness
 ├─Quick
 │    ├─Reliability=good
 │    ├─Reliability=bad, Emphaty=satisfy
 │    └─Reliability=bad, Emphaty=less satisfy, Assurance= convince
 └─Slow
      ├─Emphaty=satisfy, Reliability=good, Tangibles=less comfortable, Assurance= convince
      ├─Emphaty=satisfy, Reliability=good, Tangibles=comfortable
      ├─Emphaty=satisfy, Reliability=bad, Assurance=convince
      └─Emphaty=less satisfy, Assurance=convince, Tangibles=comfortable, Reliability=good
```
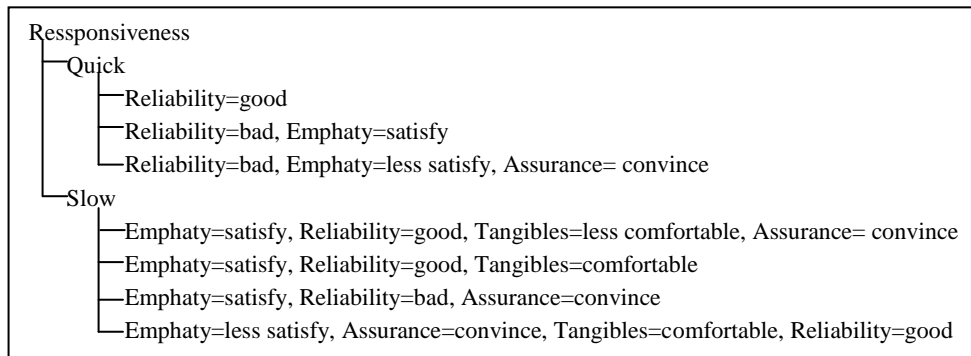
Figure 7: Loyal Customer Factor

7. ANALYSIS

From the research of service quality uses Servqual method in ID3 Algorithm; show that loyal customer can be shown in the Table 2.

Table 2. ID3 Algorithm result table

Reliability	Responsiveness	Assurance	Emphaty	Tangibles
Good	Quick	Convince	Satisfy	Comfortable
Good	Quick	Convince	Satisfy	Less Comfortable
Good	Quick	Convince	Less Satisfy	Comfortable
Good	Quick	Convince	Kurang	Less Comfortable
Good	Quick	Less Convince	Satisfy	Comfortable
Good	Quick	Less Convince	Satisfy	Less Comfortable
Good	Quick	Less Convince	Kurang	Comfortable
Good	Quick	Less Convince	Kurang	Less Comfortable
Good	Slow	Convince	Satisfy	Comfortable
Good	Slow	Convince	Satisfy	Less Comfortable
Good	Slow	Convince	Kurang	Comfortable
Good	Slow	Less Convince	Satisfy	Comfortable
Bad	Quick	Convince	Satisfy	Comfortable
Bad	Quick	Convince	Satisfy	Less Comfortable
Bad	Quick	Convince	Kurang	Comfortable
Bad	Quick	Convince	Kurang	Less Comfortable
Bad	Quick	Less Convince	Satisfy	Comfortable
Bad	Quick	Less Convince	Satisfy	Less Comfortable
Bad	Slow	Convince	Satisfy	Comfortable
Bad	Slow	Convince	Satisfy	Less Comfortable

From the all the variable Reliability, Responsiveness, Assurance, Empathy, and Tangibles can be conclude that:

- If they have quick responses and good reliability, the other variable will not be influenced. But if they has bad reliability, the other factor that should be focus is Empathy, should satisfy the customer. If the empathy is not satisfying enough, Assurance factor will play the main role and should convince the customer.

- If the response is slow, the other factor which should be focus on is empathy and Assurance or empathy and Tangibles. Empathy should satisfy the customer. Assurance should assure the customer and tangibles should make the customer feel comfortable in purchasing or transacting. While Reliability doesn't have a big influence if the responsiveness to the customer is slow.

8. CONCLUSION

Decision tree in ID3 Algorithm and WEKA 3.7.7 can solve some problems in understanding the customer loyalty factor. Both are easy to understand especially to analyze the loyalty factor. The main factor of this analysis is responsiveness. It is

related to the employee capability to help the customer and give a quick response to the customer. Responsiveness is a dynamic dimension. It is influence the employee behavior, like Clear Information, Service Speed, Service accuracy, Willingness to help, and Readiness to respond.

9. REFERENCES

[1] D. Lavanya, and K. U. Rani, "Performance Evaluation of Decision Tree Classifiers on Medical Datasets", (IJCA) International Journal of Computer Applications, Vol. 26 No. 4, July 2011.

[2] D. L. Gupta, A. K. Malviya, and S. Singh, "Performance Analysis of Classification Tree Learning Algorithms", (IJCA) International Journal of Computer Applications, Vol. 55 No. 6, October 2012.

[3] D. S. Kumar, N. Ananthi, and M. Devi, "An Approach to Automation Selection of Decision Tree based on Training Data Set", (IJCA) International Journal of Computer Applications, Vol. 64 No. 21, February 2013.

[4] Firdian Endy, Surachman, and Santoso Budi, 2012, Servqual Application Method and Six Sigma to Analise PT. PLN (Persero) Service Quality Network Service Unit (UPJ) Dinoyo Malang, Science and Engineering, Vol.13 No.3 September 2012, ISSN: 1412-7237.

[5] H. Monem, A. R. C. Hussin, and R. Sharifian, "Comprehensive Prioritized Perspectives of CRM System for Implementation in the Developing Countries Hospitals", (IJCA) International Journal of Computer Applications, Vol. 50 No. 20, July 2012.

[6] H. S. Soliman, "Customer Relationship Management and Its Relationship to the Marketing Performance", International Journal of Business and Social Science, Vol. 2 No. 10, June 2011.

[7] J. Munusamy, S. Chelliah, and H. W. Mun, "Service Quality Delivery and Its Impact on Customer Satisfaction in the Banking Sector in Malaysia", International Journal of Innovation Management and Technology, Vol. 1, No. 4, October 2010, ISSN: 2010-0248.

[8] K. Nandhini, and S. Saranya, "ID3 Classifier for Pupils' Status Prediction", (IJCA) International Journal of Computer Applications, Vol. 57, November 2011.

[9] Nino, 2012, Service Quality Infulence Analysis and Relationship Quality to Customer Loyalty on AHASS Tamtama Motor in Kediri Region, REVITALISATION Management Journal, Vol.1, No.2, September 2012.

[10] N. Malpathak, S. Shinde, and R.George, "Cloud Based CRM Application", (IJCA) International Journal of Computer Applications, April 2012, ISSN: 0975-8887.

[11] R. A. Soeini, B. Jafari, and M. Abdollahzadeh, "CRM Performance Measurement Process", (IJES) International Journal of Emerging Sciences, March 2012, ISSN: 2222-4254.

[12] T. Smitha, and V. Sundaram, "Case Study On High Dimensional Data Analysis Using Decision Tree Model", (IJCSI) International Journal of Computer Science Issues, Vol. 9, Issue 3, No. 3, May 2012, ISSN: 1694-0814.