International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Web Application for Copier Machine Rental System

Study of System Design, Data Mining Techniques of Classification and Predication Apriori Algorithm for Mining Frequent itemset for Efficient Inventory Management

Shreya Diwate[1], Dhanashree Jain[2], Trupti Bargaje[3], Rajesh Lomte[4]

[1, 2, 3]Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
Professor, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

Abstract: ***In copier machine rental business, it is necessary to predict quantity of spare parts of the copier machines for better inventory management. If the relationship between corrupted spare parts can be identified, it will be helpful to the business inventory management and for machine repairing employees also to know which parts they should keep with them before going to the customer's place. . In this paper we are discussing Apriori algorithm for recommendation and three classification algorithms-, Naïve Bayes algorithm, Logistic regression, Decision Tree algorithm. We are considering a common 3 month machine dataset and analyzing each classification algorithm for its performance and accuracy and selecting best algorithm of them.***

Keywords: Naive Bayes, Decision Tree, Logistic regression, Classification, Apriori algorithm, prediction, analysis

1. Introduction

In the recent days, we have seen the growing awareness amongst people and the shift in their attitude towards automation over manual work, in order to manage the human resources, customer services and of course to increase the profitability of the business. This goes with the business of Rental Services.

Manual management of everything is not only tedious and highly labor intensive, but it is also exposed to human errors leading to inaccuracy and inefficiency, overall affecting the business. Thus, to gain more control over the business, a web application is proposed.

This paper will discuss about the background knowledge, existing system, newly proposed system design, and the concerned algorithms.

2. Existing System

The Rental Service considered here is concerned with Copier Machines. The copier machines, to lower the upfront costs and avoid hassle of resale/disposal, are preferred to use on lease rather than purchasing. It also avoids the problem of obsolescence.

The firms managing these business caaries their operations as follows:

a) When a customer wants to buy a machine on lease, he can contact the manager, enquire and make the contract.
b) Once machines are installed, the the customer pays the monthly bill as:
   Total = B + (N' * Cost_per_unit)
   where, B = Base_rent
   N'= N – 10000
   N= Total copies printed
c) At the end of month, the customer has to send the reading to manger by call/message.
d) The manager sends the bill made in Microsoft Excel back to customer.
e) Complaints, if any, has to be submitted by call/message.
f) The manager has to view all those without failure. If anything is missed then it can lead to customer dissatisfaction.

All of the data needed to to meet above purposes is stored on the registers manually (about customers, employees and inventory).

Problems with Existing System

- Current system includes recording data of all the details regarding various activities of user and entails a lot of paper work.
- As it provides text based interface, it is not as user friendly as Graphical User Interface.
- Retrieving particular data is very hard.
- The papers may get lost. In that case, all the data will be lost.
- Possibility of human errors can affect the quaity and efficiency.

As evident above, the whole process is manual, needs so much personal attention, thus becomes hectic. Thus, we need a system that is more reliable, flexible and efficient that will remove maximum disadvantages of existing system.

3. Proposed System

To eliminate the issue in existing system we proposed a web application which will enable the manager to handle the application remotely from any machine having following functions:

Propose system feature

a) Secured login
b) Database management(Customer, Employee, Inventory, Payment, Maintenance)
c) Automatic bill generation
d) Automatic payment reminder
e) Complaint management

The customer will also be able to:

a) Enter the readings
b) View transaction history
c) View balance
d) View more products
e) Register complaints

4. System Architecture

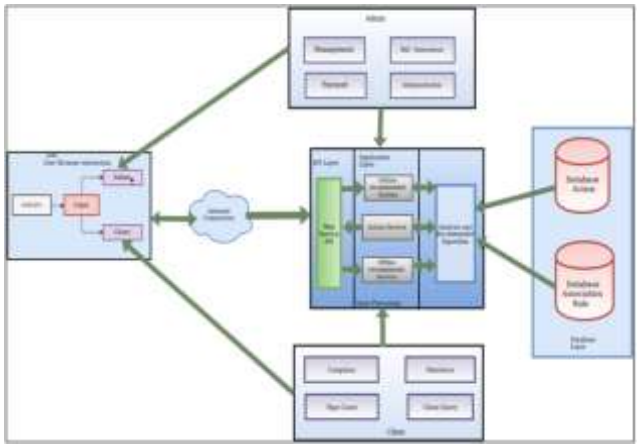


Figure 1: System Architecture

5. Classification Algorithm

A) Inventory Dataset

id	type	brand	costprice	income	maintance	class
1	Mono copiers	Sharp	147000	15634	7667	profit
2	Colour copiers	Xerox	120000	4576	6546	Loss
3	Multifunctional copiers	Konica	44000	3765	5644	Loss
4	Analog copiers	Ricoh.	300000	13441	6456	profit
5	Digital copiers	Minolta	120000	4343	5654	loss
6	Desktop copiers	Canon	14000	13424	5656	profit
7	Network copiers	Brother	2000000	5678	3453	profit
8	A3 photocopiers	Xerox	42000	24324	5435	profit
1	Mono copiers	Sharp	147000	23455	3545	profit
9	High volume photocopi	Canon	225000	3345	5454	loss
8	A3 photocopiers	Brother	42000	7985	3445	profit
1	Mono copiers	Xerox	147000	74345	3455	profit
2	Colour copiers	Canon	120000	12474	3445	profit
3	Multifunctional copiers	Minolta	44000	4000	6745	loss
4	Analog copiers	Brother	300000	5675	2345	profit
7	Network copiers	Sharp	2000000	3534	6564	loss
8	A3 photocopiers	Canon	42000	75678	6545	profit
6	Desktop copiers	Konica	14000	3522	6546	profit
8	A3 photocopiers	Minolta	42000	56544	5645	profit

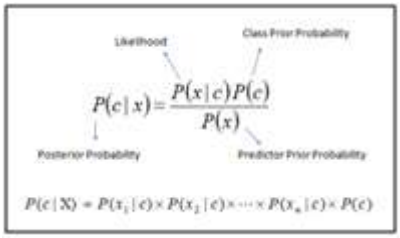
Figure 2: Maintenance Dataset

B) Algorithms

i) Naive Bayes

The Naive Bayesian classifier is based on Bayes' theorem with the independence assumptions between predictors. This method is used to find a class patterns. Naive Bayes classifier assumes that the presence of or absence of a particular feature of a class is unrelated to the presence or absence of another feature.

Bayes' theorem provides a way of measuring the posterior probability, p(c\x),from P(c),P(x) and P(x|c).Naive Bayes classifier assumes that the effect of the value of a predictors. This assumption is called condition independance.



Where,

- P(c|x) is tehe posterior probability of class(target) given predictor (attributes).
- P(c) is the prior probability of class.
- p(x|c) is te likehood which is the probability of predictor given class.
- p(x) is the prior probability of predictor.
- Mathematically Bayes theorem gives the relationship between the probabilities of c and x ,P(c) and P(x)and conditional probabilities of c given x and x given c, namely P(c|x) and p(x|c).

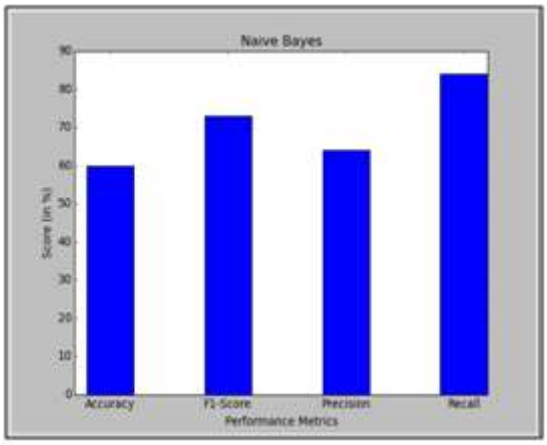


Figure 3: Naive Bayes results

ii) Logistic Regression

Logistic regression is a classification algorithm. It is a statistical method for analyzing a dataset in which there are one or more independent variables which determine an outcome. The outcome is measured with help a dichotomous variable. The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a *logit transformation* of the probability of presence of the characteristic of interest:

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

where p is the probability of presence of the characteristic of interest.

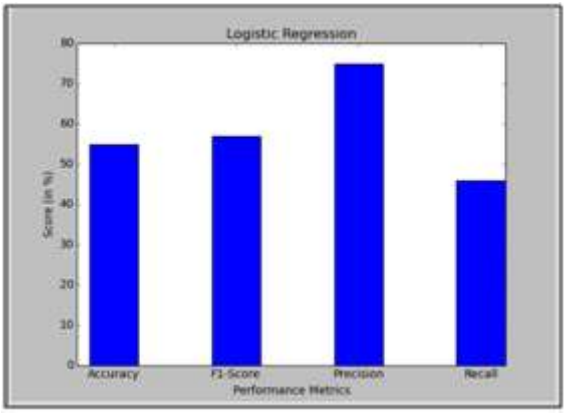


Figure 4: Logistic Regression results

iii) Decision Tree

Decision tee is a tree data structure which is used to builds classification or regression models in the form of tree structure. It breaks down a dataset into smaller and smaller subsets and simultaneously an associated decision tree is incrementally developed. The decision tree classification tree structure that include a root node, branches and leaf node. Every internal node defines test on an attributes, each branch denotes the result of test and leaf node holds class labels. The topmost element is known as root node.

The Decision Tree algorithm is based on conditional probabilities. decision trees generate rules. A rule is a conditional statement that can easily be understood by humans and used within a database to identify records.

Metrics used in decision tree (ID3)

a) **Entropy:** a measure of the impurity in a collection of training examples
   $S(x) = - \sum p(x) \log p(x)$
   p is the proportion of examples in set S for x.
b) **Information gain:** a measure of the effectiveness of an attribute in classifying the training data.
   $Gain(S,A)=Entropy(S)-\sum\frac{Sv}{S}Entropy(Sv)$

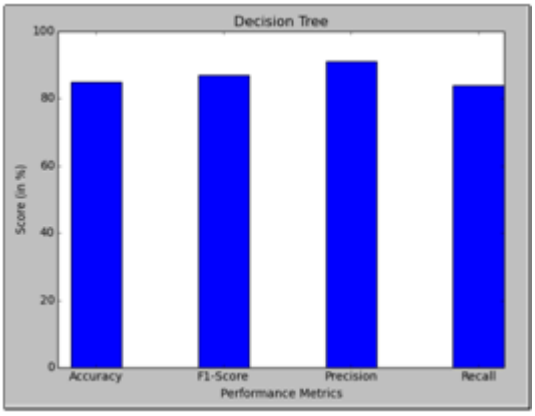


Figure 5: Decision Tree results

6. Comparison

i) Confusion matrice

Naive Bayes	Decision Tree	Logistic Regession
[1 6]	[6 1]	[5 2]
[2 11]	[2 11]	[7 6]

ii) Comparison of algorithms

Name of Algorithm	Accuracy	Precision	F Score	R call
Naive Bayes	60%	65	71	87
Decision Tree	87%	90	88	85
Logistic Regression	54%	71	56	44

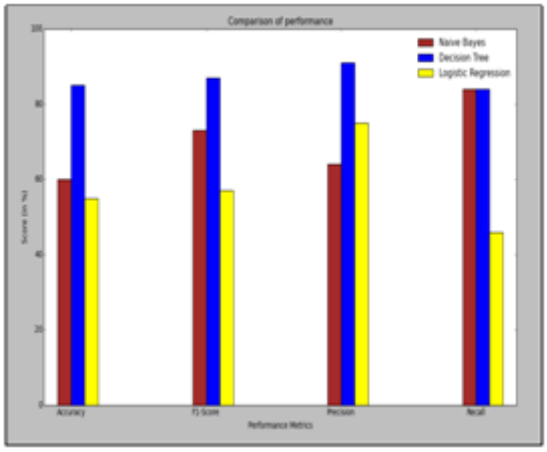


Figure 6: Comparitive Representation of Classifiction Agorithms

7. Prediction Algorithm

Apriori Algoritm

Finding the hidden relationships in large datasets is known as association rule learning or association analysis. The problem is, finding different useful and frequent combinations of items can be a time-consuming task and prohibitively expensive in terms of computing power. By using frequent item sets and association rules, retailers have a much better understanding of their customers. We first need to find the frequent itemsets, and then we can find the association rules.

The Apriori algorithm principle says that if an itemset is frequent, then all of its subsets are frequent. This means that if {0,1} is frequent, then {0} and {1} have to be frequent. The rule turned around says that if an itemset is infrequent, then its supersets are also infrequent.

Example: *Association rules can be found using Apriori Algorithm for identifying relationship between corrupted spare pars of a copier machine or printing machine.*

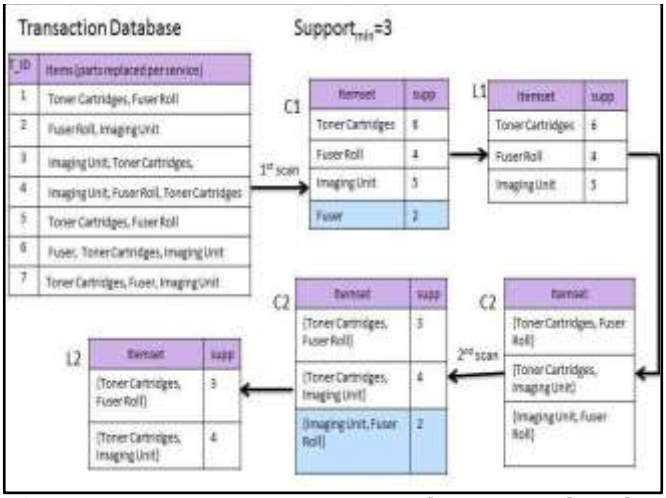


Figure 7: Frequent Itemset Generation by Apriori Algorithm

8. Conclusion

Hence through this algorithmic survey we conclude that Decision Tree is best among the three classification algorithm because it gives 87% accuracy and it takes less time to classify the data than the other algorithms. This algorithm

will definitely help to classify the machines data into whether machine is in profit or loss. We choose the apriori algorithm for finding association rules which will help to provide maintenance services in efficient and effective way.

References

[1] Yuguang Huang, Lei Li, Gang Xie, "NAIVE BAYES CLASSIFICATION ALGORITHM BASED ON SMALL SAMPLE SET", Proceedings of IEEE CCIS2011.
[2] Jiao Yabing,"Research of an Improved Apriori Algorithm in Data
[3] Mining Association Rules",International Journal of Computer and Communication Engineering, Vol. 2, No. 1, January 2013.
[4] Sheng Chai 1, 3 , Jia Yang 2 , Yang Cheng ,"The Research of Improved Apriori Algorithm for Mining Association Rules"1-4244-0885-7/07/$20.00 ©2007 IEEE.
[5] Mr. Brijain R Patel, 2 Mr. Kushik K Rana,"A Survey on Decision Tree Algorithm For Classification",2014 IJEDR | Volume 2, Issue 1 | ISSN: 2321-9939.
[6] Jaloree, Shaielesh, A. Rajput and Sanjeev Gaur "Decision tree approach to build a model for water quality" Binary Journal for Data mining & Networking 4.1 (2014): 25-28.
[7] lugendra Dongre,Gend Lal Prajapati,s. V. Tokekar,"The Role of Apriori Algorithm for Finding the Association Rules in Data Mining",2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT).
[8] Vijayalakshmi Sampath, Andrew Flagel, Carolina Figueroa, "A LOGISTIC REGRESSION MODEL TO PREDICT FRESHMEN ENROLLMENTS",Paper SD-016.
[9] "Youssef Bassil,A Comparative Study on the Performanceof the Top DBMS Systems",Vol. 1, No. 1, Pages. 20-31, February 2012.
[10] PunamKumari and 2 Rainu Nandal,"A Research Paper OnWebsite Development OptimizationUsing Xampp/PHP",Volume 8, No. 5, May-June 2017,ISSN No. 0976-5697.
[11] Bayu Waspodo, Qurrotul Aini and Syamsuri Nur,"Development Of Car Rental Management Information System",1st International Conference on Information Systems For Business Competitiveness (ICISBC) 2011.