Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

Applying Data Mining Tool to Study Consumer Perception Towards E-Banking

Computer Science

KEYWORDS : Data Mining, WEKA tool, e-banking, banking sector, classification, Clustering (Simple K Means), etc.

Sulochana Patil — Assistant Professor, Krishna Institute Of Computer Application, And Management, Wathar (Karad), Maharashtra, Indiad

Dr. Sudhakar D. Bhoite — Asso. Prof. M.Phil. Department, Ch. Shahu Institute Of Business, Education And Research, Kolhapur, Maharashtra, India

ABSTRACT

Various banks in India are now becoming more market- oriented and customer-friendly.The banking sector has started realizing the need of the techniques like data mining to analyze data from the huge size databases. It is also importantto understand the various environmental, economic and socio-cultural factors that influence buying behavior or acceptance and denial of specific services by the consumers.This paper highlights applications of data mining tool WEKA to analyze the obstacles in adoption of e-banking services.It will demonstrate the reasons of denial of e-banking facility by consumers from rural and semi urban areas in sangli district by using two different data mining techniques in WEKA: Classification, Clustering (Simple K Means).

INTRODUCTION :

The quality of customer service in banks has three dimensions - One is the accessibility of banking services, second, availability of a range of liability & asset, products and services and the third most important is the human element, which processes and delivers the service to the customers (M.D. Mallya, 2010)[1]. With the induction of technology in banks, there is gradual multi-dimensional improvement in the range of customer service.

Nowadays banks are benefitedwith virtual banking through a variety of more exciting, innovative and technology-based bank products like banking at home/office, Internet banking, mobile banking and so on, that can be summed of as e-banking. A lot need is to create confidence in the minds of customers about the benefits and security of these e-banking services.

DATA MINING :

In its simplest form, data mining automates the detection of relevant patterns in a database, using defined approaches and algorithms to look into current and historical data that can then be analyzed to predict future trends (Heikki, Mannila, 1996)[2]. Because data mining tools predict future trends and behaviors by reading through databases for hidden patterns, they allow organizations to make proactive, knowledge-driven decisions and answer questions that were previously too time-consuming to resolve.

Data mining functionalities are characterization and discrimination,mining frequent patterns, association, correlation, classification andprediction, cluster analysis, outlier analysis and evolution analysis(Han J. and Kamber M, 2000) [3]. Two of the major data miningtechniques are classification and clustering. Classification is a datamining (machine learning) technique used to predict group membershipfor data instances. CLUSTERING is a data mining techniqueto group the similar data into a cluster and dissimilar datainto different clusters. we are using WEKA data mining tool forclassification and clustering to investigate the acceptance of e-bankingfacility by bank customers in sangli district with a samplesize of 300 customers from public, private and co-operative banks.

WEKA :

WEKA is a landmark system in the history of the data mining and machine learning research communities, because it is the only toolkit that has gained such widespread adoption and survived for an extended period of time (Sapna Jain, M Afshar Aalam and M N Doja) [4]. It is a collection ofstate-of-the-art machine learning algorithms and data preprocessing tools written in Java, developed at the University of Waikato New Zealand. It is free software that runs on almost any platform and is available under the GNU General Public License. It has a wide range of applications in various data mining techniques. It provides extensive support for the entire process of experimental data mining, including preparing the input data, evaluating learning schemes statistically, and visualizing the input data and the result of learning. The WEKA workbench includes methods for the main data mining problems: regression, classification, clustering, association rule mining, and attribute selection. It can be used in either of the following two interfaces

- Command Line Interface (CLI)
- Graphical User Interface (GUI)

This research paper will demonstrate "The scenario of e-banking facility denial by bank consumers in sangli district by following two data mining techniques using WEKA: 1) Classification2) Clustering (Simple K Means)

By creating a classification tree (a decision tree), the data can bemined to determine the denial of e-banking by a person. Possible nodes on the tree would be not-aware, no-operating-knlg, feeling_unsecure, xtra-servic-charge, unefficient. These attributes of reasons can be used against the decision tree to determine the likelihood of a person rejecting e-banking facility.

The data used in this paper is Bank Data available in AttributeRelation File Format (ARFF) with following fields & 80 records.
@relation **reason-not-using-e-banking**
@attribute **Not-aware** {y,n}
@attribute **no-operating-knlg** {y,n}
@attribute **feeling_unsecure** {y,n}
@attribute **xtra-servic-charge** {y,n}
@attribute **unefficient** {y,n}

The denial_reasons.arff is loaded into WEKA and will look like –

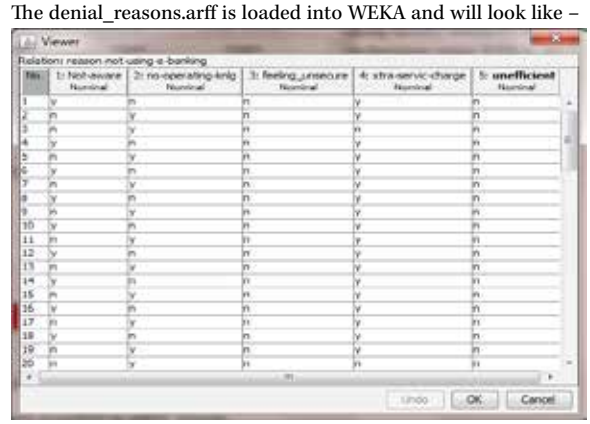

Fig. 1 - Bank Data available in ARFF format.

The Instances distribution is visualized with denial_reasons attribute as class is follows-

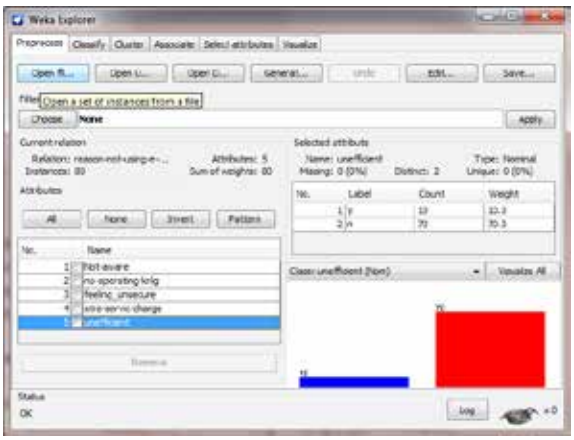

Fig. 2 - WEKA Explorer with denial_reasons class

CLASSIFICATION OF DATA:

Classification is used to classify each item in a set of data into one of predefined set of classes or groups. In tree classification j48 is an implementation of C4.5 release 8, a standard algorithm that is widely used for practical machine learning (Quinlan 1993)[5]. This implementation produces decision tree models. This algorithm works by forming pruned and unpruned partial decision trees.

The bank data for this article is classified with weka.classifiers.trees.J48 -C 0.25 -M 2 pruned method and got the classifier output as follows-

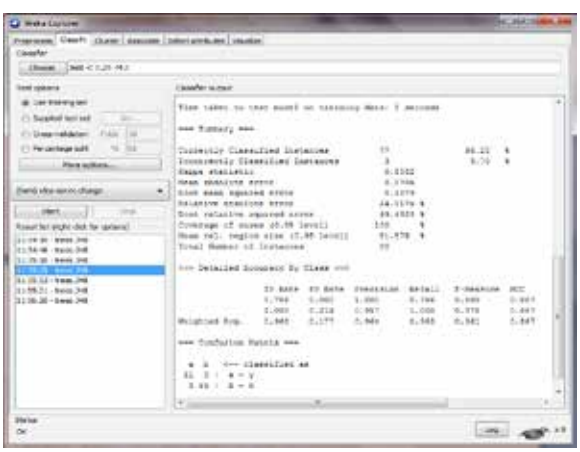

Fig. 3 - weka.classifiers.trees.J48 -C 0.25 -M 2 pruned method (training set)

Listing of Output from WEKA's classification model (above window)

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: reason-not-using-e-banking
Instances: 80
Attributes: 5
Not-aware
no-operating-knlg
feeling_unsecure
xtra-servic-charge
unefficient

Test mode: evaluate on training data

=== Classifier model (full training set) ===

J48 pruned tree
---
xtra-servic-charge = y: n (51.0/3.0)
xtra-servic-charge = n
| no-operating-knlg = y: n (12.0)
| no-operating-knlg = n
| | Not-aware = y: n (6.0)
| | Not-aware = n: y (11.0)

Number of Leaves : 4
Size of the tree : 7
Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	77	96.25 %
Incorrectly Classified Instances	3	3.75 %
Kappa statistic	0.8582	
Mean absolute error	0.0706	
Root mean squared error	0.1879	
Relative absolute error	24.0176 %	
Root relative squared error	49.4323 %	
Coverage of cases (0.95 level)	100 %	
Mean rel. region size (0.95 level)	81.875 %	
Total Number of Instances	80	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.786	0.000	1.000	0.786	0.880	0.867	0.922	0.834	y
	1.000	0.214	0.957	1.000	0.978	0.867	0.922	0.968	n
Weighted Avg.	0.963	0.177	0.964	0.963	0.961	0.867	0.922	0.945	

=== Confusion Matrix ===

a	b	<-- classified as
11	3	a = y
0	66	b = n

What do these numbers mean-

Time taken to build model: 0 seconds

Correctly Classified Instances	77	96.25 %
Incorrectly Classified Instances	3	3.75 %

Confusion matrix is another important aspect to be considered, from this matrix predictions can be made.

Confusion Matrix

a	b	<-- classified as
11	3	a = y
0	66	b = n

Above 11 instances are correctly classified for the class value Yes. And below 66 instances are correctly classified instances for the class value No. Remaining elements represents the incorrectly classified instances Yes and No respectively.

Based on our accuracy rate of 96.257 %, we can say that this is a pretty good model to analyze reasons for consumer denial of e-banking facility.

A **tree Visualization** can be seen on the model just created as follows -

Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

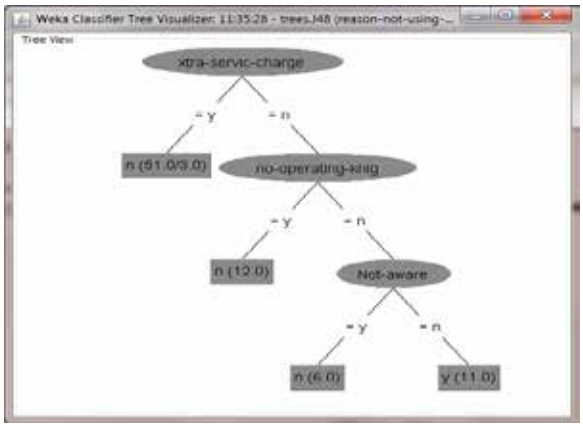

Figure.5- Classification Tree Visualization

CLUSTERING :

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Clustering is also called as data segmentation. Clustering can also be used as outlier detection, where outliers may be more interesting than common cases. Many clustering algorithms exist in WEKA. The performance of K-Means algorithm produces quality clusters when using huge dataset and is better than Hierarchical Clustering algorithm(Bharat Chaudhari, Manan Parikh,2012)[6].

Considering this, we have used K-Means algorithm to quickly determine patterns in the data.

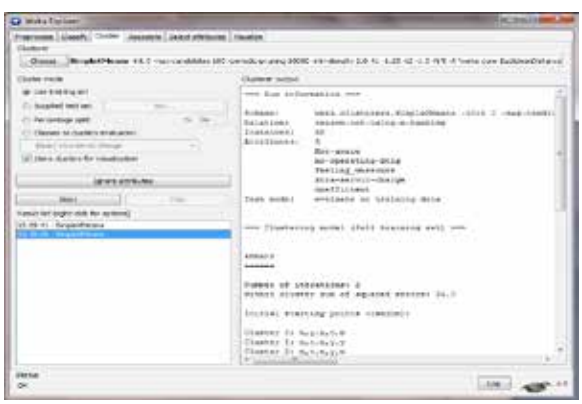

Fig. 5 Clustering by Simple K means algorithm (5 clusters)

Listing of Cluster Output with 5 clusters :

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 5 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: reason-not-using-e-banking
Instances: 80
Attributes: 5
Not-aware
no-operating-knlg
feeling_unsecure
xtra-servic-charge
unefficient
Test mode: evaluate on training data
=== Clustering model (full training set) === kMeans =====
Number of iterations: 2
Within cluster sum of squared errors: 24.0
Initial starting points (random):
Cluster 0: n,y,n,n,n
Cluster 1: n,n,n,y,y
Cluster 2: n,n,n,y,n
Cluster 3: n,n,y,n,n
Cluster 4: n,y,n,y,n

Missing values globally replaced with mean/mode
Final cluster centroids:

Attribute	Full Data (80.0)	Cluster# 0 (18.0)	1 (10.0)	2 (32.0)	3 (11.0)	4 (9.0)
Not-aware	n	n	n	n	n	n
no-operating-knlg	n	y	n	n	n	y
feeling_unsecure	n	n	n	n	y	n
xtra-servic-charge	y	n	y	y	n	y
unefficient	n	n	y	n	n	n

Time taken to build model (full training data) : 0 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 18 (23%)
1 10 (13%)
2 32 (40%)
3 11 (14%)
4 9 (11%)

Description of Clusters-

Cluster 0 - This group represents to the respondents who are not using e-banking services because of having no or less operating knowledge of e-banking services. There are 18 (23%) out of 80 respondent belongs to this group.

Cluster 1 - This group represents to the respondents who are not interested in e-banking services because they are not ready to pay extra service charges levied by banks for usage of these services. The respondents from this group also think that e-banking services are inefficient. There are 10 (13%) out of 80 respondents belong to this group.

Cluster 2 - This group represents to only those respondents who are not using e-banking services because they are not ready to pay extra service charges levied by banks for usage of these services. There are 32 (40%) out of 80 respondents belong to this group

Cluster 3 - This group represents to the respondents who are not using e-banking services because they feel that these services are unsecure. There are 11 (14%) out of 80 respondent belongs to this group.

Cluster 4 - This group represents to the respondents who are not interested in e-banking services because they are not ready to pay extra service charges levied by banks for usage of these services. The respondents from this group also have no operating knowledge of e-banking services. There are 9 (11%) out of 80 respondents belong to this group.

One thing that is clear from above cluster distribution is that, respondents of extra service charges are heavily distributed among 5 clusters.

CONCLUSION :

The primary objective of the study was to study consumer denial of online banking in Sangli district in the light of the technology acceptance. The study proposed that online banking denial can be modeled with five variables derived as from the Not-aware, no-operating-knlg, feeling_unsecure, xtra-servic-charge, unefficient.

The results of the analysis conducted on the five factors indicate that amount of information on online banking were found to be

the most influential factors explaining the use of online banking services. This finding refers to the fact that consumers deny online banking for the extra service charges levied by banks for usage of online banking services in comparison to other factors considered in the specified study.

Finally, it proves that WEKA is a significant step in the transfer of machine learning technology into the workplace.

REFERENCES

1. M.D. Mallya, "ROLE OF TECHNOLOGY IN ENHANCING QUALITY OF CUSTOMER SERVICE IN BANKS" Published by S. S. Bhandare for the Forum of Free Enterprise, Penins~da House, 2nd Floor, 235, Dr. D. N. Road, Mumbai 400001, and Printed by S. V. Limaye at lndia Printing Works, lndia Printing House, 42 G. D. Arnbedkar Marg, Wadala, Mumbai 400 031.
2. Heikki, Mannila. 1996. Data mining: machine learning, statistics, and databases, IEEE, http://en.wikipedia.org/wiki/Data_mining.
3. Han J. and Kamber M.: "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, San Francisco, 2000.
4. Sapna Jain, M Afshar Aalam and M N Doja, "K-means clustering using weka interface", Proceedings of the 4th NationalConference; INDIACom-2010.
5. **Quinlan, J.R. (1993) C4.5: Programs for machine learning. Morgan Kaufmann, San Mateo, CA.**
6. **Bharat Chaudhari, Manan Parikh, "A Comparative Study of clustering algorithms Using weka tools", Web Site: www.ijaiem.org Email: editor@ijaiem.org, editorijaiem@gmail.com Volume 1, Issue 2, October 2012 ISSN 2319 - 4847 Page 158.**