International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Simulation Model for Predicting Banking Customers Satisfaction Using Chi-Square Test & Naïve Bayes Classifier

Mahesh Mahajan
Assistant Professor,
SVKM's Institute of Technology, Dhule(M.S.)
mahesh.mahajan@svkm.ac.in

Reeta Singh
Assistant Professor,
ASM's IMCOST, Thane(M.S.)
reeta.singh27july@gmail.com

Abstract:

Today, it is necessary for all Public, Private and Government bank to find their banking customers satisfaction ratio depending on feedback provided by the customers. These research paper is useful for finding Personal Characteristics and Banking Behavior of Customers. The study uses primary data which was collected from 550 responders. The responders are taken from different profession and different age group. The bank can offer different kind of services like ATM, Net banking, Mobile banking and so on. The objective of the paper is to find the result of customer satisfaction. The result is found by Non-probabilistic technique chi- Square test and probabilistic technique Naïve Bayes Classifier. The chi-square test (symbolized as $\chi 2$) is basically done by analyzing the data of the agreed upon method for the d ifference between variables. It is usually a comparison of two statistical data.
Naive Bayes algorithm is based on Bayes theorem and is used to solve classification problems. The result of study indicates both positive and negative influence on banking customer behavior.

Keywords

Simulation, chi-square test, Naïve Bayes Classifier, demographic segmentation, customer behaviors.

1. Introduction

Customer profiling helps to target customers for the banking security products. India has a very diverse system of banks. These are:

- Public
- Private
- Co-operative
- Development Bank/Financial institutions

The study was conducted with 550 retail bank customers.

2. THE BANKING INDUSTRY IN INDIA

The Indian banking industry faced a multitude challenges. There was a visible lack of product expertise. The banks were traditionally focused on limited range of products, primarily for corporate clients.

The banks had a relatively high unit cost of delivery given small transaction sizes. There was a limited use of technology; across both, customer-facing and internal functions within the banks.

The way ahead for the banks is to scale up to the international standards in terms of service, use of technology and other competitive advantages to survive in the competitive environment. One major factor for the banks is to retain the old customers and to acquire new customers across board; be it corporate clients, or individual retail customers.

The banks need to achieve this via consistent and systematic marketing and sales efforts. The banks already have a lot of data with them in the form of retail customer data; transaction data and many studies done in the past few years. The data could be mined for retail customer profiling to turn existing customers into ambassadors and trying to acquire new customers based on existing data. The data can be used to understand further retail customer needs, gaps in the market and for product research. This could help the banks in projecting newer services, locations and different avenues of earnings.

3. Research Methods

For the research, the following research method has planned.

Instrument Design: Questionnaire form is designed to measure the variables and characteristics of interest.

Sample size: 550 sample data is collected for the study.

Sampling Technique: Non-Probability and Probability sampling technique

Statistical tools applied:

- Chi square test - To find the association between two variables
- Cluster analysis – For segmenting and classifying the data
- Naïve Bayes Classifier- For Classification

4. Demographics

Demographic segmentation divides a business's customers by

- age,
- gender,
- marital status,
- family type/size,
- family life cycle,
- job, etc.

5. Customer Profiling

A behavioral and satisfaction profile models of the customer depends on the customer's actions and is usually derived from customer's transactional data.

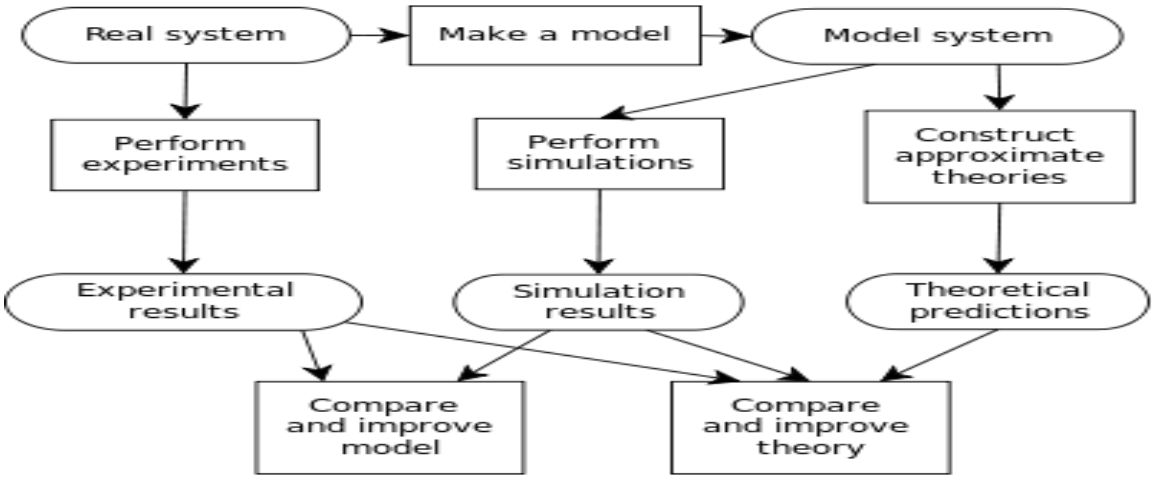
6. Customer Relationship Management

In an age of fierce competition, the customer is God. Simulation can be useful in all the three phases of a customer relationship cycle: Customer Acquisition, Increasing value of the customer and Customer retention.

7. Simulation

Simulation is the imitation of a real-world process or process over time. Simulating the behavior of something before building a model. This model represents the main features, behavior and functions of a selected physical or abstract problem or process. While the model represents the system itself, the simulation represents the functioning of the system over time.

8. Computer Simulation



Computer simulations use mathematical models to model the behavior of a system. Computer simulations can be small computers that can run almost instantly on a small device, or large programs that run for hours or days on a cluster of networked computers. Computer simulations model events on a scale far beyond what is possible (or even imaginable) using mathematical models.

9. Simulation in Banking Sector

The goal of the bank simulation is to facilitate understanding of the personnel characteristics and banking customer behavior and their satisfaction index.

10. Chi-Squared Test ($\chi 2$)

Chi-square test is a statistical hypothesis test in which the sample of the test statistic is a chi-square distribution when the null hypothesis is true.

In evaluation practice, observations are divided into two specific groups and there are some or no assumptions that give the probability of an observation being in the same category. It is use to check that, the hypothesis is true or not , given the observations.

11. Findings and suggestions:

For finding and suggestions purpose we apply chi-square test and Naïve Bayes classifier.

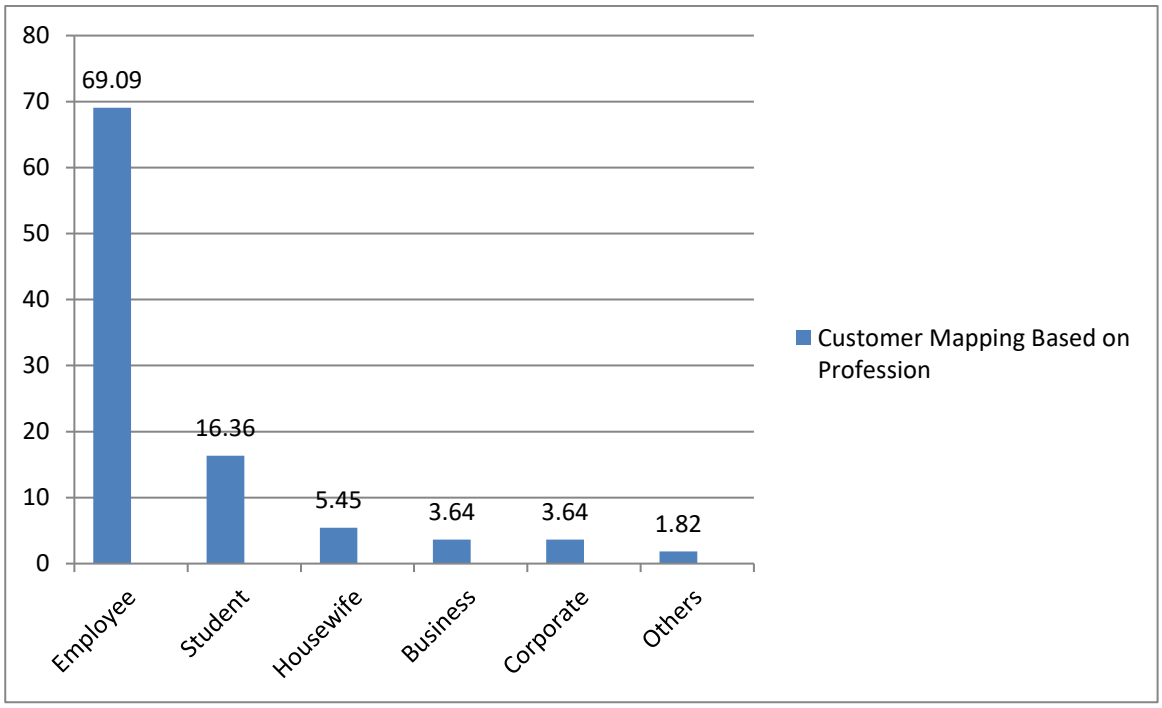
11.1 Chi-square($\chi 2$) Test:

The outcome depends upon the mapping of customers based on professions, mapping of customers based on their age group, banking services use by the customers and details of the Customers mode of transaction with the bank.

The study on Retail Customer Profiling, Simulation Techniques to Strategies Sales & Marketing in Indian Banks has paved way to a series of findings as discussed below. The study was conducted with 550 retail bank customers.

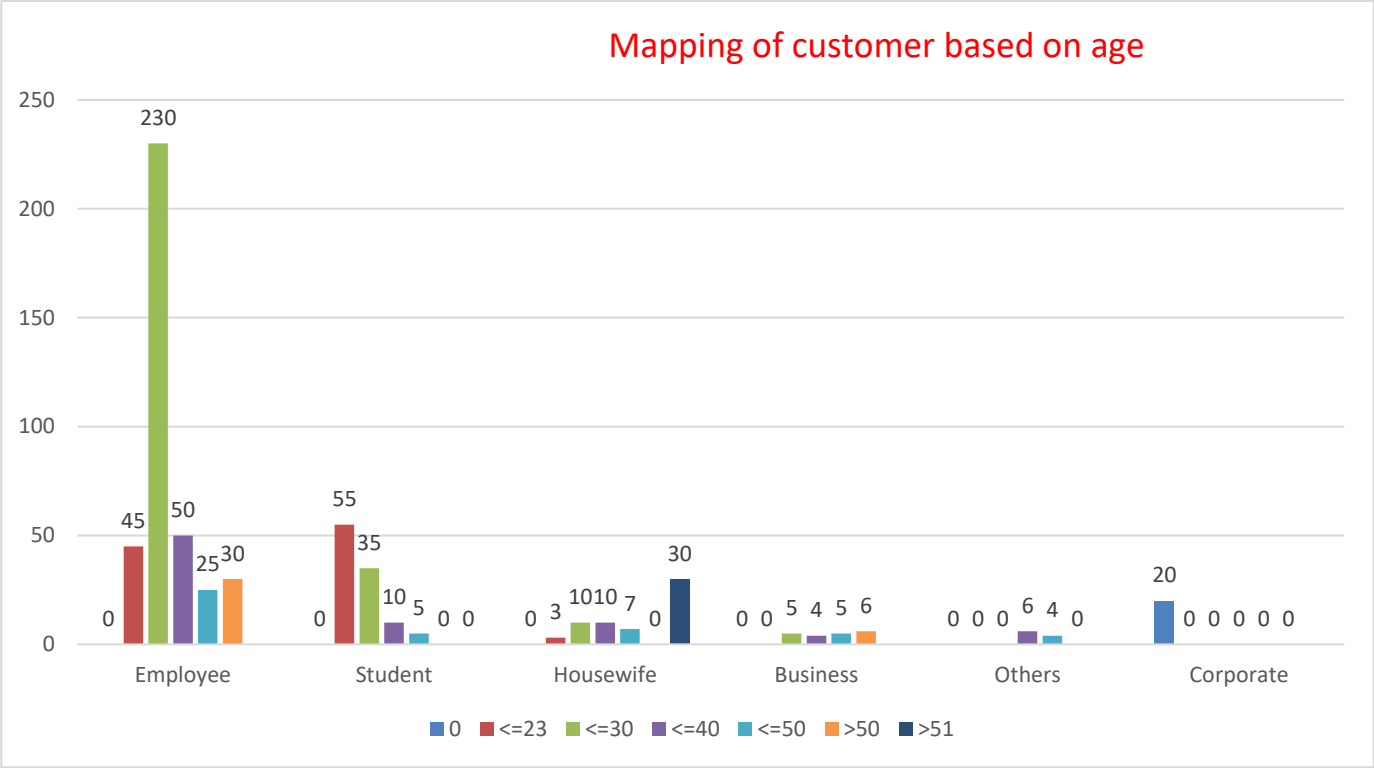
1. The following table shows the mapping of customers based on professions.

Employee	Student	Housewife	Business	Corporate	Others	Total
380	90	30	20	20	10	**550**
69.09%	16.36%	5.45%	3.64%	3.64%	1.82%	**100%**



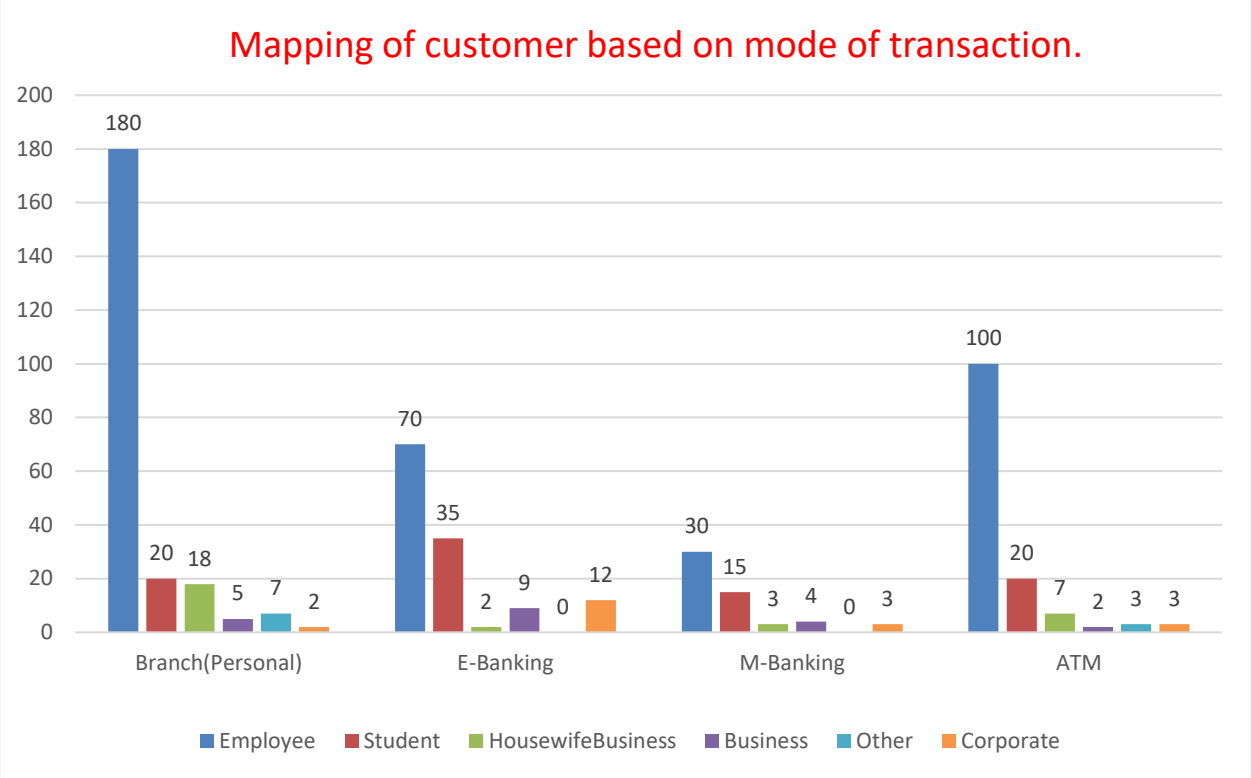
2. The following table shows the mapping of customer based on age.

Age	Employee	Student	Housewife	Business	Others	Corporate	Total
=18	0	0	0	0	0	20	20
<=23	45	55	03	0	0	0	103
<=30	230	35	10	05	0	0	280
<=40	50	0	10	04	06	0	70
<=50	25	0	07	05	04	0	41
>50	30	0	0	06	0	0	36
Total	380	90	30	20	10	20	**550**



3. The study has promoted to look into the detail of the Customers mode of transaction with the bank. The following table shows the mapping of customer based on mode of transaction.

Channel Used	Employee	Student	Housewife	Business	Other	Corporate	Total	%
Branch (Personal)	180	20	18	05	07	02	232	42.18
E-Banking	70	35	02	09	0	12	128	23.27
M-Banking	30	15	03	04	0	03	55	10.00
ATM	100	20	07	02	03	03	135	24.55
Total	380	90	30	20	10	20	**550**	



4. Depending on the services use by the customers, we define following null and alternative hypothesis.

Under the null hypothesis we define,

H_0 = denotes that, the customers are satisfied.

The alternative hypothesis is,

H_1= Customers are not satisfied.

Customers Types	Number of Customers		Total
	Satisfied with bank services	Not satisfied with bank services	
Employees	275	105	380
Students	55	35	90
Housewife	21	9	30
Business	12	8	20
Corporate	11	9	20
Others	8	2	10
Total	382	168	550

Here frequencies are arranged in the form of a 6×2 contingency table. Hence the degree of freedom is (6-1) × (2-1) =5×1=5.

Under the null hypothesis of independence, expected frequencies (Ei) are calculated as-

Customers Types	Number of Customers		Total
	Satisfied with bank services	Not satisfied with bank services	
Employees	382*380/550=263.94	380-263.94=116.06	380
Students	382*90/550=62.50	90-62.50=27.50	90
Housewife	382*30/550=20.84	30-20.84=9.16	30
Business	382*20/550=13.89	20-13.89=6.11	20
Corporate	382*20/550=13.89	20-13.89=6.11	20
Others	382*10/550=6.94	10-6.94=3.06	10
Total	382	168	550

The following table shows the calculations for Chi-Square-

Customers Types	Satisfied/Not Satisfied	Observed Frequencies (Fi)	Expected Frequencies (Ei)	Fi-Ei	$(Fi-Ei)^2$	$(Fi-Ei)^2 / Ei$
Employees	Satisfied	275	263.94	11.06	122.3236	0.4634523
Students		55	62.5	-7.5	56.25	0.9
Housewife		21	20.84	0.16	0.0256	0.00122841
Business		12	13.89	-1.89	3.5721	0.25717063
Corporate		11	13.89	-2.89	8.3521	0.6013031
Others		8	6.94	1.06	1.1236	0.16190202
Employees	Not Satisfied	105	116.06	-11.06	122.3236	1.05396864
Students		35	27.5	7.5	56.25	2.04545455
Housewife		9	9.16	-0.16	0.0256	0.00279476
Business		8	6.11	1.89	3.5721	0.58463175
Corporate		9	6.11	2.89	8.3521	1.36695581
Others		2	3.06	-1.06	1.1236	0.36718954
		$\sum Fi = 550$	$\sum Ei = 550$			$\sum(Fi-Ei)^2 / Ei = 7.80605149$

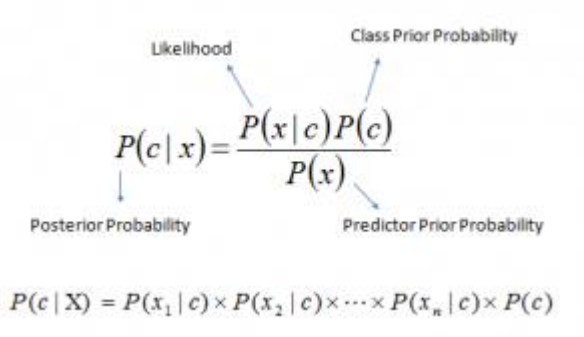
The tabulated value of $\chi 2$ is 11.070 at 5% level of significance.

Conclusion: we observed that calculated value of $\chi 2$ < tabulated value, at 5% level of significance and hence, we accept the null hypothesis. Hence we conclude that the customers are satisfied with banking services.

11.2 Naïve Bayes Classifier:

It is most popular supervised machine learning algorithm and it is a classification system based on Bayes' theorem with the assumption of independence of predictor variables.

According to Bayes theorem ,



Where,

$P(c|x)$: posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).

$P(c)$: prior probability of *class*.

$P(x|c)$: likelihood which is the probability of the *predictor* given *class*.

$P(x)$: prior probability of the *predictor*.

Consider the following frequency table:

Customers Types	Number of Customers		Total
	Satisfied with bank services	Not satisfied with bank services	
Employees	275	105	380
Students	55	35	90
Housewife	21	9	30
Business	12	8	20
Corporate	11	9	20
Others	8	2	10
Total	382	168	550

Customers Types	Number of Customers		Total	Probability
	Satisfied with bank services(Yes)	Not satisfied with bank services(No)		
Employees	275	105	380	380/550=0.69
Students	55	35	90	90/550=0.16
Housewife	21	9	30	30/550=0.05
Business	12	8	20	20/550=0.04
Corporate	11	9	20	20/550=0.04
Others	8	2	10	10/550=0.02
Total	382	168	550	
Probability	382/550=0.69	168/550=0.31		

(Likelihood Table 1)

Customers Types	Number of Customers		Posterior Probability for Yes	Posterior Probability for No
	Satisfied with bank services(Yes)	Not satisfied with bank services(No)		
Employees	275	105	275/382=0.72	105/168=0.63
Students	55	35	55/382=0.14	35/168=0.21
Housewife	21	9	21/382=0.05	9/168=0.05
Business	12	8	12/382=0.04	8/168=0.05
Corporate	11	9	11/382=0.03	9/168=0.05
Others	8	2	8/382=0.02	2/168=0.01
Total	382	168		

(Likelihood Table 2)

1. Now suppose we want to calculate the probability of customer satisfaction when the customer is Employee.

Probability of Yes:

P (Yes | Employee) = P (Employee | Yes) P(Yes) / P (Employee)----------(1)

P(Employee) = 380/550 = 0.69

P(Yes)= 382/550 = 0.69

Posterior Probability: (Employee |Yes) = 275/382 = 0.72

Recall equation (1)

P (Yes |Employee) = 0.72 * 0.69 / 0.69 = 0.72(Higher)

Probability of No:

P(No | Employee) = P(Employee| No) P(No) / P (Employee) .....(2)

Calculate Prior Probabilities:

P(Employee) = 380/550 = 0.69

P(No)= 168/550=0.31

Calculate Posterior Probabilities:

P(Employee |No) = 105/168=0.63

Recall equation equation (2)

P (No | Employee) = 0.63 * 0.31 / 0.69 = 0.28

The probability of a 'Yes' class is higher. So, we can conclude that Employees are satisfied with banking services.

2. Now suppose we want to calculate the probability of customer satisfaction when the customer is Student.

Probability of Yes:

P (Yes | Student) = P (Student | Yes) *P(Yes) / P (Student)*----------*(*1)

P(Student) = 90/550 = 0.16

P(Yes)= 382/550 = 0.69

Posterior Probability: (Student |Yes) = 55/382 = 0.14

Recall equation (1)

P (Yes |Student) = 0.14 * 0.69 / 0.16 = 0.60(Higher)

Probability of No:

P(No | Student) = P(Student| No) P(No) / P (Student) .....(2)

Calculate Prior Probabilities:

P(Student) = 90/550 = 0.16

P(No)= 168/550=0.31

Calculate Posterior Probabilities:

P(Student |No) = 35/168=0.21

Recall equation (2)

P (No | Student) = 0.21 * 0.31 / 0.16 = 0.40

The probability of a 'Yes' class is higher. So, we can conclude that Students are satisfied with banking services.

3. Now suppose we want to calculate the probability of customer satisfaction when the customer is Housewife.

Probability of Yes:

P (Yes | *Housewife*) = P (*Housewife* | Yes) *P(Yes) / P (Housewife)*----------*(*1)

P(Housewife) = 30/550 = 0.05

P(Yes)= 382/550 = 0.69

Posterior Probability: (Housewife |Yes) = 21/382 = 0.05

Recall equation (1)

P (Yes | Housewife) = 0.05 * 0.69 / 0.05 = 0.69(Higher)

Probability of No:

P(No | Housewife) = P(Housewife | No) P(No) / P (Housewife) .....(2)

Calculate Prior Probabilities:

P(Housewife) = 30/550 = 0.05

P(No)= 168/550=0.31

Calculate Posterior Probabilities:

P(Housewife |No) =9/168=0.05

Recall equation (2)

P (No | Housewife) = 0.05 * 0.31 / 0.05 = 0.31

The probability of a 'Yes' class is higher. So, we can conclude that Housewife's are satisfied with banking services.

4. Now suppose we want to calculate the probability of customer satisfaction when the customer is Business Type.

Probability of Yes:

P (Yes | *Business*) = P (*Business* | Yes) *P(Yes) / P (Business)*----------(1)

P(Business) = 20/550 = 0.04

P(Yes)= 382/550 = 0.69

Posterior Probability: (Business |Yes) =12/382 = 0.04

Recall equation (1)

P (Yes | Business) = 0.04 * 0.69 / 0.04 = 0.69(Higher)

Probability of No:

P(No | Business*)* = *P(*Business | *No)* P(No) / P (Business) .....(2)

Calculate Prior Probabilities:

P(Business) = 20/550 = 0.04

P(No)= 168/550=0.31

Calculate Posterior Probabilities:

P(Business |No) =8/168=0.05

Recall equation (2)

P (No | Business) = 0.05 * 0.31 / 0.05 = 0.31

The probability of a 'Yes' class is higher. So, we can conclude that Businessman's are satisfied with banking services.

5. Now suppose we want to calculate the probability of customer satisfaction when the customer is Corporate Type.

Probability of Yes:

P (Yes | *Corporate*) = P (*Corporate* | Yes) *P(Yes) / P (Corporate)*----------(1)

P(Corporate) = 20/550 = 0.04

P(Yes)= 382/550 = 0.69

Posterior Probability: (Corporate |Yes) =11/382 = 0.03

Recall equation (1)

P (Yes | Corporate) = 0.04 * 0.69 / 0.03 = 0.69(Higher)

Probability of No:

P(No | Corporate*)* = *P(*Corporate | *No)* P(No) / P (Corporate) .....(2)

Calculate Prior Probabilities:

P(Corporate) = 20/550 = 0.04

P(No)= 168/550=0.31

Calculate Posterior Probabilities:

P(Corporate |No) =9/168=0.05

Recall equation (2)

P (No | Corporate) = 0.05 * 0.31 / 0.05 = 0.31

The probability of a 'Yes' class is higher. So, we can conclude that Corporates are satisfied with banking services.

6. Now suppose we want to calculate the probability of customer satisfaction when the customer is Other.

Probability of Yes:

P (Yes | *Other*) = P (*Other* | Yes) *P(Yes) / P (Other)*----------(1)

P(Other) = 10/550 = 0.02

P(Yes)= 382/550 = 0.69

Posterior Probability: (Other |Yes) =8/382 = 0.02

Recall equation (1)

P (Yes | Other) = 0.02 * 0.69 / 0.02 = 0.69(Higher)

Probability of No:

P(No | Other*)* = *P(*Other | *No)* P(No) / P (Other) .....(2)

Calculate Prior Probabilities:

P(Other) = 10/550 = 0.02

P(No)= 168/550=0.31

Calculate Posterior Probabilities:

P(Other |No) =2/168=0.01

Recall equation (2)

P (No | Other) = 0.01 * 0.31 / 0.02 = 0.31

The probability of a 'Yes' class is higher.

Using both Chi-Square Test & Naïve Bayes' Classifier we conclude that, banking customers are satisfied .

12. Conclusion and Future Scope

In banking industry customers are focal point. The customer's satisfaction is totally based on the services provided by the banks and different technology use by the bank. The way of providing the services based on computer technology and mobile technology is greatly accepted by the customers. We conduct the study in one particular city and it can extend to other cities.

References

1. S.S.Rai, Anil T. Gaikwad, R.V. Kulkarni, "A Research Paper on Simulation Model for Teaching and Learning Process in Higher Education.", International Journal of Advanced Computer Research (ISSN(print): 2249-7277 ISSN(online): 2277-7970), Volume -4 Number-2 Issue-15 June-2014.

2. Rajeshri Lanjewar and Prakash Yadav, "Uderstanding of Customer Profiling and Segmentation Using K-Means Clustering Method for Raipur Sahkari Dugdh Sangh Milk Products", International Journal of Research in Computer and Communication Technology,Vol 2, Issue 3, March-2013.

3. Dr. Manasa Nagabhushanam, "A Study on Customer Service Quality of Banks in India", Analyz Research Solutions Pvt. Ltd. Bangalore.

4. Learn Naive Bayes Algorithm : Naive Bayes Classifier Examples (analyticsvidhya.com)

5. Learn Naive Bayes Algorithm : Naive Bayes Classifier Examples (analyticsvidhya.com)