

Predicting Customer's Satisfaction (Dissatisfaction) Using Logistic Regression

Adarsh Anand¹, Gunjan Bansal^{2*}

Department of Operational Research
University of Delhi, Delhi 110007, India
¹adarsh.anand86@gmail.com, ²gunjan.1512@gmail.com
**Corresponding author*

(Received June 21, 2016; Accepted July 2, 2016)

Abstract

Customer satisfaction is a metric of how products and services offered by companies meet customer expectations. This performance indicator assists companies in managing and monitoring their business effectively. Firms thus need reliable and representative measure to know the customer satisfaction. In the present work, we provide a predictive model to identify customer's satisfaction (dissatisfaction) with the firm's offerings. For the analysis, "mobile phone" has been used as a product and 11 related decision making variables have been taken as independent variables. Due to the dichotomous (i.e. satisfaction/ dissatisfaction) nature of the dependent variable, a powerful tool among multivariate techniques i.e. Logistic Regression has been applied for the validation. Further, Receiver Operating Characteristic (ROC) curve has been plotted which displays the degree to which the prediction agrees with the data graphically. The analysis has been done on data collected from students of University of Delhi, Delhi.

Keywords-customer's dissatisfaction, customer's satisfaction, logistic regression, multivariate technique, receiver operating characteristic (ROC) curve.

1. Introduction

"Marketing is the delivery of customer satisfaction at a profit." (Kotler and Armstrong, 2014). A market is a place which is made of two significant entities i) "The producers" and ii) "The customers". Producers explore new opportunities, analyse them and choose the best opportunity which supports firms and companies to come up with the new or improvised products. An opportunity is basically the current needs or requirements of the customers which has been generated due to dynamic environment. Hence, it becomes imperative for the firms to observe on regular basis that how much expectations of the customers have been met from their current offering. For this purpose, companies spend millions of money, conduct many surveys, observe different markets, and monitor their competitors etc. And, on the basis of observations and feedback, companies improvise their existing products in terms of product quality, service quality, design etc. or turn up with the totally new offering.

Hence, the overall performance of firms is directly or indirectly dependent on the customers' satisfaction. Though it is said that customer's retention is more significant than customers' satisfaction. But, customers' satisfaction plays a major role in retaining the customers i.e. if a customer is satisfied from firms' product then there are more chances of continuing of the same product or buying any other product from the same company (brand) rather than a customer who is dissatisfied. It is stated that, overall satisfaction is a key indicator of firm's success because it impacts on behaviour and economic aspects of the firms (Anderson et al., 1994). Research has shown that higher satisfaction generates customers' loyalty which leads to customers' retention (Hennig and Hansen, 2000). Also, customers' satisfaction has a strong and positive impact on

willingness to pay more. Many programs have been implemented for measuring and improving customer satisfaction since it becomes one of the most important focus of the corporate strategies.

Adopting immediate strategy of proper and suitable actions against any complaints results in increasing loyalty and positive words of mouth which generates customers' satisfaction (Brown and Beltrani, 1989). An article written by Scott Smith in 2012, has written seven different types of expectations or experiences which a customer may receive from any product or services. These seven expectations are termed as: i) Explicit expectations ii) Implicit Expectations iii) Static Performance Expectations iv) Dynamic Performance Expectations v) Technological Expectations vi) Interpersonal Expectations vii) Situational Expectations. He also advised that to get the accurate insights of any research based on customer satisfaction, one should incorporate these factors (Smith, 2012).

For building a healthy marketing relationship three important standpoints or variables have been mentioned i.e. customer satisfaction (or service quality), trust and commitment. Customer satisfaction and service quality sometime have been considered as a similar construct that measures the difference between the expectations which a customer expects from an offering (or from its quality) and the actual performance of the product. If the difference between the actual performance and expectation are higher than it leads to dissatisfaction. Trust is believed as a mutual construct; "if firm ask for customer trust, they should also trust their customer in return" (cows, 1996). In marketing relationship trust subsist when product and their services are reliable and have immense integrity. Commitment implies to build a long term business relationship of faith that provides benefits to each other (Hennig and Hansen, 2000).

Evaluation of customer satisfaction is one good factor that helps in many decision making strategies. (Kapur et al., 2014) but, predicting customers' behaviour, their levels of satisfaction/dissatisfaction always been a difficult challenge for companies. How various statistical techniques can be useful for analysing customer satisfaction have been very well explained in (Allen and Rao, 2000). In this study, we have developed a predictive model that determine whether the customer is satisfied (or dissatisfied) from the offering given by the company. To accomplish this study, we have chosen a very well-known and widely used statistical technique called Logistic Regression, detail description of this technique have been explained in section 2. To examine how well a model has been fitted graphically, concepts of ROC (Receiver Operating characteristic) curve have been discussed in section 3. Section 4 describes the research methodology with numerical illustration in subsection 4.1 that is based on the particular product i.e. "Mobile Phones". Section 5 explicated the interpretation and significant insights of this study. Managerial implication given in section 6, explain how this study can be useful for any company. Importance of this study has been concluded in section 7.

2. Logistic Regressions

Utilization of various statistical procedures by markets, firm and companies and its stakeholders has increased rapidly in such dynamic and competitive world. These techniques direct and help them to tackle real life problems with less erroneous, correct and appropriate results. In this case study, one of the most suitable techniques called logistic regression has been applied.

Logistic regression is a specialized form of a very common statistical method but widely used technique called multiple regressions. The name of this technique derives from the log- it transformation with the dependent variable. The dependent variable follows a binomial

distribution and behave as a dichotomous (or binary) in nature. That is only two values are possible either 0 or 1. Mathematically, dependent variable can be expressed as $y = \pi(x_1, x_2, \dots, x_q) + \varepsilon$, where $\pi(x_1, x_2, \dots, x_q)$ logit function and ε is the error term. If $y = 1$, then $\varepsilon = 1 - \pi(x_1, x_2, \dots, x_q)$ with the probability $\pi(x_1, x_2, \dots, x_q)$ and if $y = 0$ then $\varepsilon = \pi(x_1, x_2, \dots, x_q)$ with probability $1 - \pi(x_1, x_2, \dots, x_q)$. So ε has a distribution with mean zero and variance equal to $\pi(x_1, x_2, \dots, x_q)(1 - \pi(x_1, x_2, \dots, x_q))$, i.e. the conditional distribution of binary response variable follows a binary distribution with the probability given by the conditional mean, $\pi(x_1, x_2, \dots, x_q)$. So, for modeling the expected value of the response directly as a linear function of independent variables, an appropriate transformation is defined. This transformation is the *logistic* or *logit* function of π leading to the model:-

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q. \quad (1)$$

The logit of a probability is simply the log of the odds of the response taking the value one. Eq. (1) can be rewritten as:-

$$\pi(x_1, x_2, \dots, x_q) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_q x_q)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_q x_q)}. \quad (2)$$

Any real value can be taken by this logit function, but the associated probability will always be lie within the required interval $[0, 1]$. In a logistic regression model, the parameter β_j associated with explanatory variable x_j is such that $\exp(\beta_j)$ is the odds that the response variable takes the value one when x_j increases by one, conditional on the other variables remaining constant. (Everitt and Hothorn, 2006; Hosmer and Lemeshow, 2000; Hair et al., 2009).

Newer and more statistically appropriate methods such as CHAID, logit, log-liner models provide more acceptable and reliable results as compare to the traditional approaches like multiple regression, discriminant analysis and AID (Automatic Interaction Detection), which has been proved by (Magidson, 1988). Many modern enterprises, in fact in such a highly competitive environment every company and firm are employing huge databases due to availability of data in massive quantity. For the purpose of data analysis, the growth of data mining techniques has been increasing rapidly. In the formulations and the solutions of various data mining problem as optimization problems have significantly used the concept of operational research techniques (Olafsson et al., 2006). Various empirical comparison have been illustrated which compare the logistic regression with other mixture of techniques of data mining such as Neural Networks, RFM (Recency, Frequency and Monetary) and CHAID (Chi-square Automatic Interaction Detection) (Kumar et al., 1995; Mccarty and Hastak, 2006). Each and every technique has both strengths and weaknesses which should be clearly understood before applying in any study. In data mining techniques, logistic regression is one of the supervised techniques that is used for linear classifying the data only when two groups are present (Witten and Frank, 2005; Elayidom, 2015).

3. Receiver Operating Characteristic (ROC) Curve

This method quantifies the accuracy of the predicted model that discriminate the predicted values between positive (true) and negative (false) cases by plotting the area under the curve (AUC). The value of the AUC varies from 0.5 (discriminating power not better than chance) to 1.0 (perfect discriminating power) and therefore it is referred as c-statistic (or concordance index). ROC curve is a graphical representation of the true positive rate (sensitivity) against the false positive rate (1-specificity), which have been plotted for different cut points. The correspondence of the sensitivity/specificity pair is represented by each point.

Sensitivity: probability that the predicted result will be true when the actual case is also true (true positive rate, expressed as a percentage)

$$\text{Sensitivity} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

Specificity: probability that the predicted value will be false when the actual case is also false (true negative rate, expressed as a percentage)

$$\text{Specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}}$$

Positive likelihood ratio: ratio between the probability of a true predicted value given the *actual value is true* and the probability of a true predicted value given the *actual value is false*,

$$\text{Positive likelihood ratio} = \frac{\text{True positive rate}}{\text{False positive rate}} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

If the AUC is close to 100% then it can be said that there is no overlap in the two distributions or perfect discrimination is presented. Therefore, the closer the ROC curve is to the upper left corner, overall accuracy of the test is higher (Hanley and McNeil, 1982; Zweig and Campbell, 1993; MedCalc, 2016).

4. Research Methodology

“The young generation are ‘addicted’ to mobile phones” an article that states how mobile phones/cell phones have become one of the basic need of the young generation (Alleyne, 2011). Therefore, this study aims to evaluate the satisfaction (or dissatisfaction) of the youth. For this purpose, 18 to 25 years of age group of 101 students of University of Delhi have been surveyed. In this study, a statistical tool SPSS version 20 has been used.

4.1. Numeric Illustration

Satisfaction of the product cannot be measured using single attribute only. Therefore, in this study we have considered 11 decision related variables which impact on the overall satisfaction of the mobile phones. These 11 exogenous variables can be defined as follows:-

- (i) **Service Quality**: How much customers are satisfied with the quality of service given by any brand (or company) after the sale of the product.

- (ii) **Value for Money:** It implies how much a customer is satisfied after spending a good amount of money over the product.
- (iii) **Trusted Brand:** How much customers are satisfied with their trusted or reliable brand.
- (iv) **Brand is centric towards its customers:** This attribute implies how much customers’ are satisfied with the reaching ability of the brand towards them.
- (v) **Visibility:** It implies how much customers’ are pleased with the clarity of the display of the handset.
- (vi) **Popularity:** It defines how much customers’ are satisfied with the popularity (or promotions) of the phone and its brand.
- (vii) **Durability:** How long the handset work and how much a customer is satisfied.
- (viii) **Different schemes:** Different proposals given by any brands like Discounts and how much you are satisfied.
- (ix) **Accessories with phones:** How much a customer is satisfied with the extra stuffs (items) provided with the product.
- (x) **Company you prefer to deal with:** How much customers prefer or satisfied to deal with the brand.
- (xi) **Phone type:** A phone type (Smart phone or a simple phone) customers are currently using.

And, overall satisfaction is considered as the endogenous variable which says how much customers is satisfied with the overall product (i.e. mobile phone). Students were asked to provide their satisfaction level for dependent and all (initial 10) independent variables on the basis of the 5 pointer scale where 5 stands for “Strongly Satisfied”, 4 means “Satisfied”, 3 means “Neither Satisfied Nor Dissatisfied”, 2 means “Dissatisfied” and 1 stands for “Strongly Dissatisfied”. Descriptive information of initial 10 variables can be seen in Fig. 1. Here it can be understood that more 50% of the students are satisfied with the individual attributes like service quality, value for money, trusted brand, brand is centric, visibility, popularity and durability.

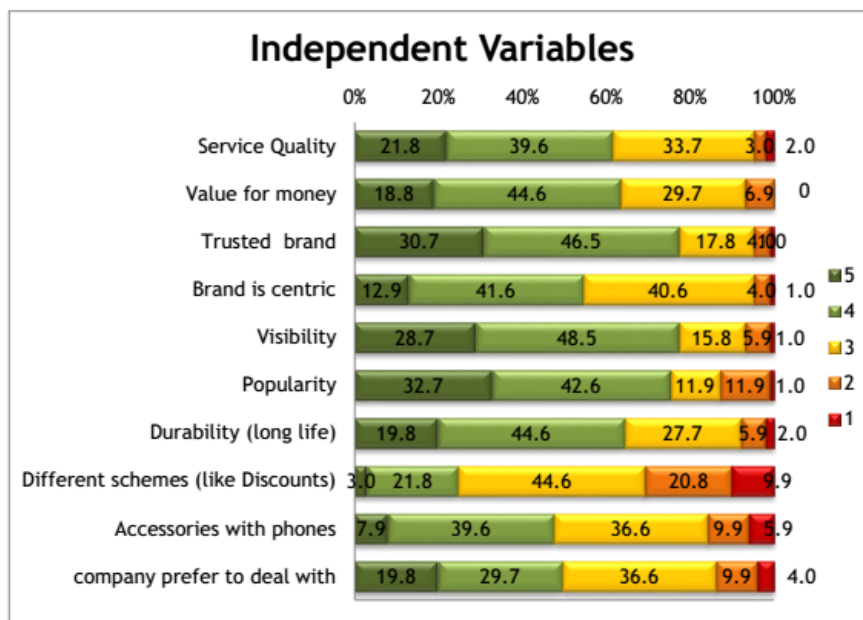


Fig.1. Descriptive information of exogenous variables

The percentage of students using Smart phone is 77.2% and 22.8% of students are still using simple phone. Graphically representation of “Phone Type” can be shown in Fig. 2.

For the purpose of simplicity and easy to understand, we have transformed these 5 pointers variables (including dependent variable as well) as dichotomous variables (or in binary form). That implies either a student can be satisfied or dissatisfied. Therefore, top two values (4 and 5) are considered as “satisfied” students and bottom three values are clubbed as “dissatisfied” students. Fig. 3 is a representation of such transformation.

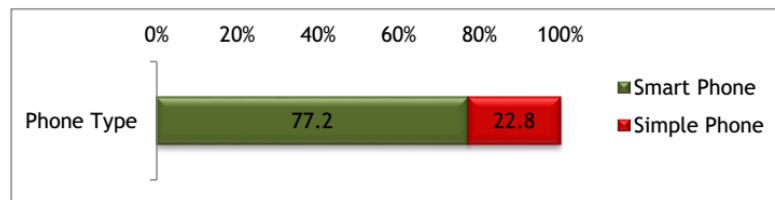


Fig. 2. Transformed information of exogenous variables

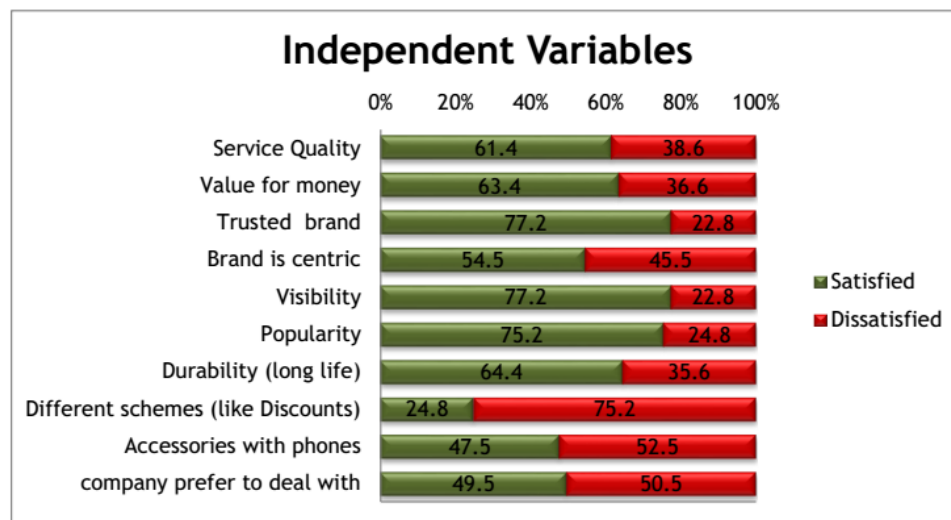


Fig. 3. Transformed information of exogenous variables

4.1.1. Correlation

Correlation is the most common statistical measure that measures the significant strength of the linear association between two or more variables. The value of pearson correlation coefficient ranges between -1 to 1. The sign indicates the direction and the magnitude of the relationship between the two variables; where absolute value of the correlation value if it is close to 1 implies that it has strong linear relationship otherwise have less related to each other (if less than 0.5) or no relation (if it is 0). Table 1 represents the relationship of all independent variables with each other.

4.1.2. Numerical on Logistic Regression

Logistic regression is a special form of multiple regressions but it differs from it in many ways. Unlike, in multiple regressions the objective is to minimize the sum of the squared difference

between the predicted and actual values of dependent variables; logistic regression employs the concept of the maximum likelihood procedure. In this procedure, the most likely estimates for the coefficients have been identified by using iterative phenomena. Hence, logistic regression identifies the maximum likelihood of an event that can occur. Stepwise (forward/ backward) estimation is the process of estimating the significant independent variables that entered sequentially according to their t-statistical value. Here, “Backward stepwise likelihood ratio” has been selected that include all candidate variables at initial step (i.e. step (0) and ended with the most significant variables (in our case it is step (8)).

4.1.2.1. Goodness-of-Fit assessing for the Estimated Model

For logistic regression model, the goodness of fit cannot be relied only on a particular measure. To assess the goodness of the model; two approaches are examined. These scenarios can be explained as below:

- (i) **-2 Log Likelihood:** In logistic regression, lower the value of -2LL implies the better fitting the model.
- (ii) **Pseudo R²:** To represent the overall model fit pseudo R² is measured as R² (or adjusted R²) is interpreted for multiple regressions. Cox and Snell R² and Nagelkerke R² are two measures considered as pseudo R². Higher values indicate the greater model fit.

Correlations										
	Service Quality	Value for money	Trusted brand	Brand is centric	Visibility	Popularity	Durability	Different schemes	Accessories	company you prefer
Service Quality	1	0.241*	0.297**	0.296**	0.151	0.252*	0.429**	0.031	-0.019	0.338**
Value For Money		1	0.322**	0.130	0.077	0.088	0.206*	0.198*	0.189	0.383**
Trusted Brand			1	0.215*	0.381**	0.400**	0.483**	0.093	0.091	0.443**
Brand Is Centric				1	0.357**	0.305**	0.316**	0.202*	0.074	0.309**
Visibility					1	0.454**	0.286**	0.093	-0.003	0.254*
Popularity						1	0.244*	0.169	0.086	0.247*
Durability							1	0.092	0.005	0.323**
Different Schemes								1	0.327**	0.166
Accessories									1	0.208*
Company You Prefer										1

Table 1. Correlations of exogenous variables

Table 2, represents the respective values of the goodness-of-fit values for -2 log likelihood, Cox and Snell R² and Nagelkerke R² for the last iteration (i.e. step 8).

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
8	52.930	0.392	0.613

Table 2. Goodness-of-fit criterions

Classification table is a cross tabulated format which provides means of assessing the predictive ability, where diagonal elements are the correct classification and off diagonals elements represents the wrong (incorrect) classifications. Table 3, represents two different classification tables i.e. 1) at initial level (step 0) when no other parameter have been considered and 2) at the last step of the iteration (step 8) when all significant parameters have been achieved.

Classification Table					
Observed			Predicted		
			Satisfaction or Dissatisfaction	Satisfied	Percentage Correct
Step 0	Satisfaction or Dissatisfaction	Dissatisfied	0	21	0
		Satisfied	0	80	100
Overall Percentage					79.2
Step 8	Satisfaction or Dissatisfaction	Dissatisfied	15	6	71.4
		Satisfied	6	74	92.5
Overall Percentage					88.1

Table 3. Classification tables

Table 4, represents the four significant variables which have been obtained among all 11 independent variables. As per this study, these four variables are most influencing attributes which highly impact on the overall satisfaction of the mobile phone. The Beta values (or logistic coefficients) can be shown in column 2. Using column 3 (wald) and column 5 (significance), wald statistic and significance of the logistic coefficients' can be analysed. Last column (Exp(B)) represents the exponentiated logistic coefficients which is the antilog of the logistic coefficients' and used for interpretation.

Variables in the Equation							
		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 8 ^a	Phone Type(1)	-2.014	1.058	3.620	1	0.057	0.133
	Service Quality(1)	1.465	0.757	3.746	1	0.053	4.327
	Value for money(1)	2.702	0.779	12.031	1	0.001	14.903
	Durability (1)	3.657	0.986	13.752	1	0.000	38.749
	Constant	1.221	0.445	7.520	1	0.006	3.392

Table 4. Significant exogenous variables for logistic regression

$$\pi(x) = \frac{e^{1.221+1.465*\text{Service Quality}+2.702*\text{Value for money}+3.657*\text{Durability}-2.014*\text{Phone_Type}}}{1 + e^{1.221+1.465*\text{Service Quality}+2.702*\text{Value for money}+3.657*\text{Durability}-2.014*\text{Phone_Type}}} \quad (3)$$

Eq. (3) represents the logistic model that can be used to predict the satisfaction/dissatisfaction for the mobile phone of any new customer. For example, suppose if a customer is having a “smart phone” and only satisfied with the “service quality” than keeping “value for money” and “durability” as zero and “smart phone” and “service quality” as one than the value of $\pi(x) = .66195$, which is greater than 0.5 (cut point) hence the customer would be considered as “satisfied” customer.

4.1.3. Numerical on ROC Curve

Table 5 represents the amount of area covered by the predictive model and using Fig. 4, graphically it can be seen that how true positive rate (specificity) is plotted against the false positive rate (1-specificity).

Area Under the Curve		
Test Result Variable(s): Satisfaction or Dissatisfaction		
Area	Std. Error	Asymptotic Sig.
0.881	0.054	0

Table 5. Area under the curve

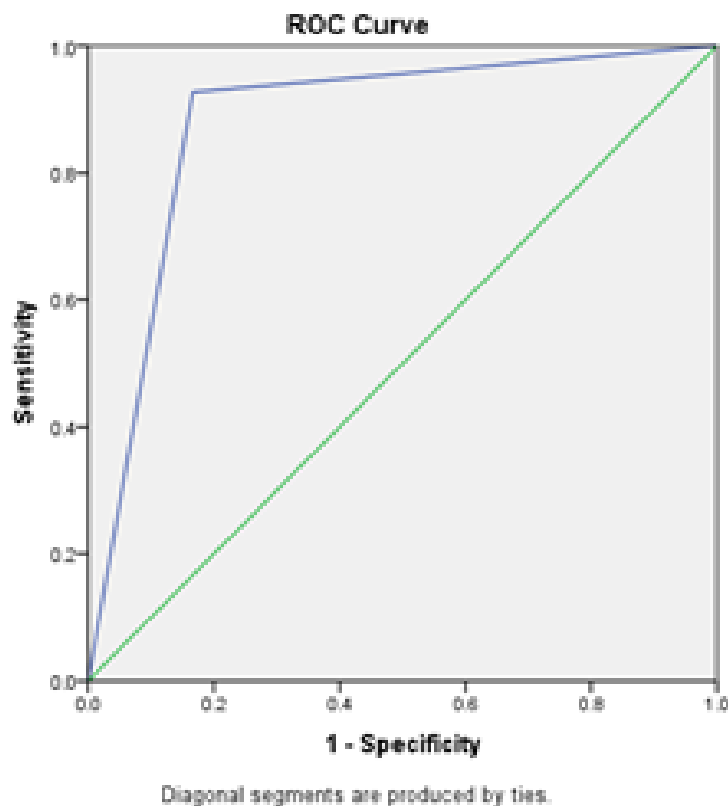


Fig. 4. ROC curve

5. Interpretation and Insightful Findings

Using Table 1, we observed the relationship between each and every independent variable has very less associated with other independent variables. Since, the variables are not highly correlated with each other it implies that they do not have multi-co-linearity. Therefore, these variables can be used for further steps (i.e. can be used for Logistic Regression). Table 2, represents the goodness of fit criterions used for logistic regression. For this particular case study, of mobile phones, satisfactory values 52.930, 0.392 and 0.613 have been obtained for $-2LL$, Cox and Snell R^2 and Nagelkerke R^2 respectively.

Table 3 i.e. “classification tables”, here it can be analysed that at the zero step, when all the customers are considered as “satisfied” as null hypothesis and no other parameter has been taken into consideration than only actual satisfied customers would be come as correct classification such that 80 students out of 101 were correctly identified. But after incorporation of the significant parameters at the 8th step, the accuracy of identifying a satisfied/dissatisfied customer has been increased by 9% i.e. from 79.2% to 88.1%. This implies that the overall satisfaction of the product depends on the satisfaction with the individual attributes of the product. For this study, type of phone, service quality, value for money and durability are those related decision variables which impact the overall satisfaction of the mobile phone.

Using Table 4, it can be interpreted that “Phone Type” has a negative impact with the satisfaction i.e. students are more satisfied with the simple mobile phones as compare to the smart phones might be the reason that with the simple phone people have less or no expectations but the students who are using smart phones have higher and more expectations which a firm (or brands) have to work on. And, other three considerable variables such as service quality, value for money and durability have positive impact with the satisfaction. It can be said that for the “durability” explanatory variable such that: durability increases the odds of satisfied respondent will remain satisfied by 38.749 times the odds of non-satisfied respondent keeping other variables constant. Similarly, other variables can be interpreted. Also, it can be seen that “durability” is the most desirable requirement among all other requirement by the customers that is if the mobile phone would work for long period, it would leads to overall satisfied customer. Table 5 and Fig. 4 represent the amount of are covered by this predictive model that 88% accuracy has been achieved.

6. Managerial Implications

Customer satisfaction is a key performance indicator for the companies; it provides a picture of how their existing customers are satisfied from their products and services. Also, it is important to analyse “why” some customers are dissatisfied at the same time so that imperative steps can be obtained. For any brand, customer satisfaction is important because of many reasons such as it helps in reducing negative word of mouth, increases intention of repurchasing, it is easier and cheaper to make existing customer more satisfied as compare to the new customers.

Therefore, this study provides a way of evaluating and predicting whether the existing/new customer is satisfied/ dissatisfied from their current offering. Also, it provides the substantial individual attributes of the mobile phones, which suggest firms on what attributes/factors they should work on more efficiently so that overall satisfaction can be increased.

7. Conclusion

This case study is an illustration that shows how the accuracy of the prediction can be improved by considering the impactful factors in evaluating and analysing the overall satisfaction of the product or service. Logistic Regression, a most profound statistical technique has been applied for developing such a model. Roc curve also have been plotted for better understanding the classification of the respondents. Hence, we have developed the predictive model based on customer satisfaction level towards the product (mobile phones).

References

- Allen, D. R., Allen, D. R., & Rao, T. R. (2000). *Analysis of customer satisfaction data: A comprehensive guide to multivariate statistical analysis in customer satisfaction, loyalty, and service quality research*. Asq Press.
- Alleyne, R. (2011). The young generation are 'addicted' to mobile phones, *The Telegraph*, <http://www.telegraph.co.uk/technology/8458786/The-young-generation-are-addicted-to-mobile-phones.html>, Accessed Date- 17-April-2016.
- Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *The Journal of Marketing*, 53-66.
- Brown, S. P., & Beltramini, R. F. (1989). Consumer complaining and word of mouth activities: field evidence. *NA-Advances in Consumer Research Volume 16*.
- Cowles, D. L., (1996), To trust or Not to Trust, background paper, *1st Internet onference on Relationship Marketing*, http://www.mcb.co.uk/services/conference_relation.mar./new_phil/backgnd2.htm, April 18th
- Elayidom, M. S. (2015). *Data mining and business intelligence*, Cengage Learning India Pvt. Ltd.
- Everitt, B. S. & Hothorn, T. (2014). *A handbook of statistical analyses using R*. CRC press.
- Hair, J. F(Jr.), Balck W. C., Babin, B. J., Anderson, R. E., Tatham R. L. (2009). *Multivariate Data Analysis*, Pearson Education Limited, Sixth Edition.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.
- Hennig-Thurau, T. (2000). *Relationship marketing: gaining competitive advantage through customer satisfaction and customer retention*. Springer Science & Business Media.
- Hosmer, D. and Lemeshow, S. (2000). *Applied logistic regression* (Wiley Series in Probability and Statistics (Vol. 2, pp. 2-4). Hoboken: Wiley-Interscience.
- Kapur, P. K., Anand, A., & Sachdeva, N. (2014). Profit estimation for a product under warranty: an evaluation based on customer's expectation and satisfaction. *International Journal of Reliability, Quality and Safety Engineering*, 21(06), 1450029.
- Kotler, P. & Armstrong, G. (2014). *Principles of marketing*, Pearson Education Limited, Fifteenth Edition.
- Kumar, A., Rao, V. R., & Soni, H. (1995). An empirical comparison of neural network and logistic regression models. *Marketing Letters*, 6(4), 251-263.
- Magidson, J. (1988). Improved statistical techniques for response modeling: Progression beyond regression. *Journal of direct marketing*, 2(4), 6-18.
- McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *Journal of business research*, 60(6), 656-662.

- MedCalc - User-friendly statistical software, (2016) <https://www.medcalc.org/manual/roc-curves.php>,
Accessed Date: 17-April-2016.
- Olafsson, S., Li, X., & Wu, S. (2008). Operations research and data mining. *European Journal of Operational Research*, 187(3), 1429-1448.
- Smith, S. (2012). *Customer Expectations: 7 Types all Exceptional Researchers Must Understand*, <http://www.qualtrics.com/blog/customer-expectations/>, Accessed Date: 17-April-2016.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 39(4), 561-577.