ANALYSIS OF CUSTOMER SATISFACTION BASED ON SERVICES PROVIDED BY AN AIRLINE COMPANY

Tanvi Saini, Narayanam Sai Lakshmi Jayani, Ishwinder Kaur

Abstract:

In airline industry, service quality, gender, class and age have an important role in influencing customer satisfaction. This paper serves to add the knowledge by improving the understanding of how these factors affect customer satisfaction for an airline company. To test the hypothesis, data was taken from Kaggle. The results of data analysis show that, in overall, that services improve customer satisfaction level. Further, we fit different models to predict the satisfaction of customers based on different factors and compare their accuracies, sensitivity and specificity. This study underlines that the service quality should be given more attention for developing customer satisfaction airlines.

Introduction:

Airline industry is highly competitive and customers are most important factor of the traveling process. Besides enhancing service quality and flight safety, customer satisfaction is the most important strategies of the airlines. In today's competitive marketing environment, service organizations mostly focus on serving customer needs. The goal is to keep up with the competition and to deliver satisfying financial benefits to account holders and other stakeholders. Customer complaints serve as a critical dimension of service quality and customer satisfaction. Complaint handling has a great affect on customer retention and the positive usage of complaint handling for service quality improvements have been extensively acknowledged by the airlines and evaluative firms.[16] Complaint management is still a focal point of study as more firms are convinced that defensive marketing is a highly cost-effective.[16] Customer complaints offer organizations with a chance to correct their mistakes, retain dissatisfied customers, and manipulate customers' future. Customers' attributions about breakdown and recovery of service then complaint satisfaction and service quality attitudes.

Studies have shown that customer dissatisfaction significantly degrades the success of service organizations in the marketplace as each dissatisfied customer on average communicates their experience to ten other individuals. As a result, successful service organizations do their best to acquire the highest level of customer satisfaction and aim to provide outstanding service for their customers. High grade customer satisfaction is a key factor to run the business, as the airline industry is very competitive and customer loyalty varies with small changes in the services. In this era of social networking, the customers get to Twitter, Facebook, Blogs or Google Reviews to review or share their experience on the Internet. Any miserable experience of a customer, which is shared becomes viral on the Internet, seriously affecting the brand or goodwill of the company. To arrest these unwanted viral trends and to avoid heavy legal compensations or penalties, the companies in the service industry are identifying a customer's satisfaction through a service rating card or survey after delivering service and compensating a dissatisfied customer upon the fault of company service. This rewards or compensation program has effectively reduced the complaints raised by a customer.

Review of Literature:

Customer satisfaction can differ from person to person and product to product. But generally if the product has at least met the needs of the consumer then it is said to be customer satisfaction. In case it fails to meet the minimum expectation then it will be turned into dissatisfaction. Tolman was the first person to use the term expectation in the context of behaviour. In general terms, expectations borrow from Tolman's expectancy theory whereby, subsequent to learning, people actualize or ward off potential consequences of their actions. Pretrial beliefs about a product that serve as standards or reference points against which product performance is judged is a commonly used definition of expectations that draws from Tolman's original conceptualization.[4] Customer satisfaction is measured in a given reference of time. So, with due respect of time even it changes so as the satisfaction level. It changes from time to time and factor to factor as it is a dynamic process. In highly involvement decisions it is very important to meet the satisfaction level. If it fails to meet the expected level then the companies will lose the customer. As there won't be any second chance. The key to provide the excellent service is in understanding the customer expectation. Expectations play a role in the formation of satisfaction and

service quality through the expectancy disconfirmation paradigm and the gap model. Satisfaction and perceived quality result from a comparison of the level of performance perceived and the level of performance expected by the consumer. Two additional components of service expectations, namely functional and technical dimensions, are found. Consumers make service evaluations based on the technical dimension of what is delivered and on the functional dimension of how, why, where, and when it is delivered.[12] A passenger, for example, will be provided with a seat (technical) passenger will interact with the cabin crew (functional).[11] There are many areas in the airlines itself where faulty processes are leading to customer complaints and displeasure. Which in return creates a negative word of mouth. Compliant handling is very important for service quality in the airline industry and it is even acknowledged by the airlines. Customers complain only when customers feel something will be done if customers know nothing will be done then customers will not only avoid traveling on the same airline but will create negative word of mouth. [9],[10]

The managers in Airlines must try to increase will and decrease should expectations or, 64 alternatively, attempt to increase will expectations and leave should expectations intact. Although the emphasis on manipulating will expectations will find dissenting opinions exists. Complaint handling is not only effective but is a very cost effective way to achieve better customer's satisfaction as well.

The Airlines have made their operations better so they were able to reduce operational costs and can give better fares with respect to other airlines. Airlines have started customer relationship programs so as to have better relations with customers so they can travel on the same airline again and again. But there were many airlines that went bankrupt because of high operational costs and low revenues. Even major airlines failed to make profits after 9/11. During the last eight years there were many airlines who ceased there operations mainly due to revenues and there were many airlines who are going in loss or hardly on break up.Due to this there were many mergers in the airline too. All this happened due to the economic conditions across the world.

Material and Method:

In order to study the customer satisfaction for Invistico Airlines, various factors like Gender, Customer Type, Age, Type of Travel, Class, Flight Distance, Seat Comfort, Departure/ Arrival time convenient, Food and Drink, Gate Location, Inflight Wifi Service, Inflight Entertainment, Online Support, Ease of Online Booking, On Board Service, Leg Room Service, Baggage, Handling, Check-In Service, Cleanliness, Online Boarding, Departure Delay in minutes, Arrival Delay in minutes have been considered. Various kinds of regression techniques and models have been used to predict the customer satisfaction.

Data

The necessary secondary data for the analysis has been obtained from kaggle

The variable type and their methods of computation have been discussed below.

Quantitative research method is used for research purposes. We employed the method of secondary data collection from Kaggle, a data science community

Statistical Techniques:

Logistic Regression: It is a statistical **model** that in its basic form uses a **logistic** function to **model** a binary dependent variable which is categorical. The coefficients (Beta values b) of the logistic regression algorithm must be estimated from your training data. We utilise the Logistic Regression classification algorithm from the GLM package in R to predict our binary dependent variable from the set of independent variables available. The best coefficients would result in a model that would predict a value very close to 1 (satisfied) for the default class and a value very close to 0 (dissatisfied) for the other class.

Setting up the null hypothesis for testing validity of regression of Satisfaction level on various factors:

Ho: Regression is invalid, i.e all

Decision Tree:

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. We use a decision tree algorithm, CART (Classification and

Regression Tree) for classifying the dependent variable based on the set of independent variables. Here the algorithm repeatedly partitions the data into multiple subsets so that the leaf nodes are homogeneous.
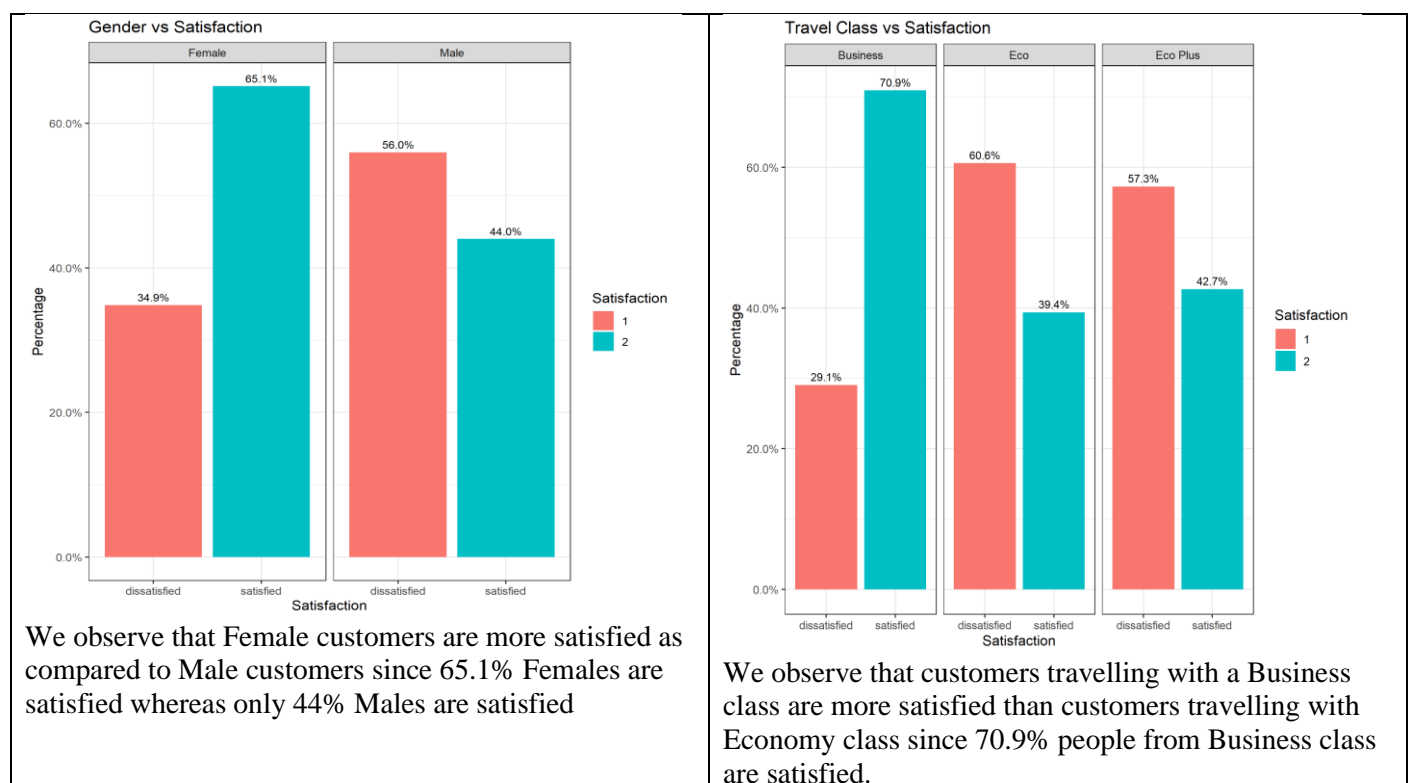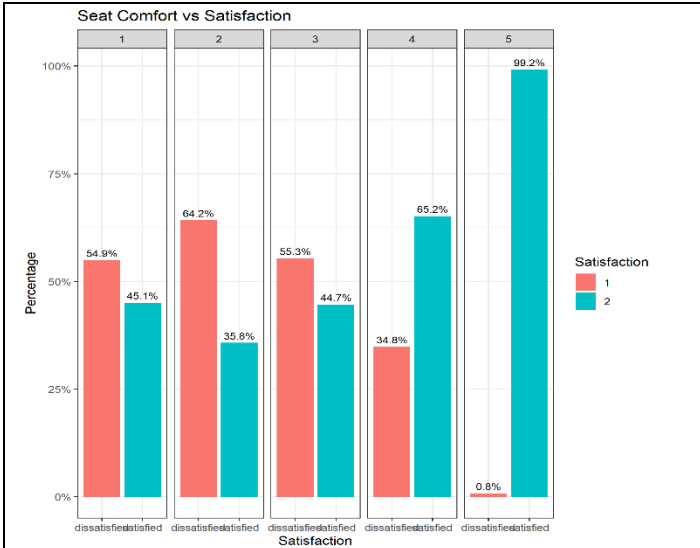
Naive Bayes - Classification Algorithm:
Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

Result and Discussion

Data Visualizations:

After observing the impact of several independent variables over the Satisfaction level, the below visualizations display a level of satisfaction at different factor levels for variables "Gender", "Travel Class", "Seat Comfort", "Inflight Entertainment", "Online Support" and "Baggage handling". These variables are considered the most significant in terms of airline customer satisfaction. Hence, we observe the following visualizations.

We observe that Female customers are more satisfied as compared to Male customers since 65.1% Females are satisfied whereas only 44% Males are satisfied

We observe that customers travelling with a Business class are more satisfied than customers travelling with Economy class since 70.9% people from Business class are satisfied.
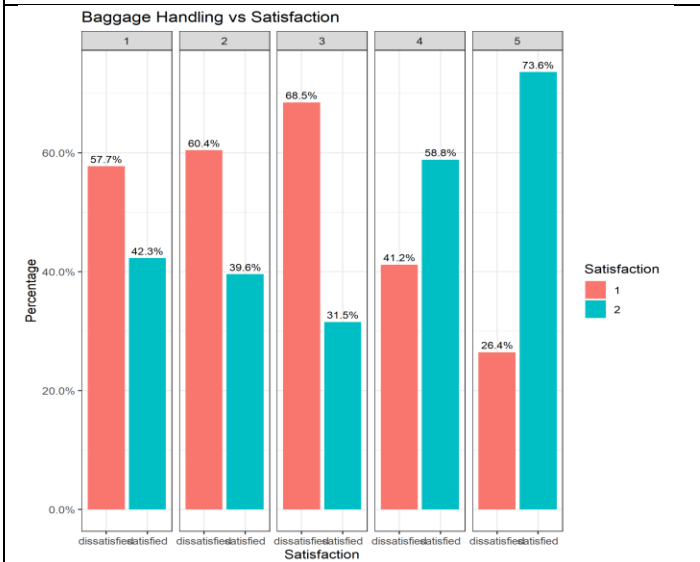
Seat Comfort vs Satisfaction

We observe that Seat Comfort has a significant effect on customer satisfaction since people with Rating 5 are 99.2% satisfied.

Online Support vs Satisfaction
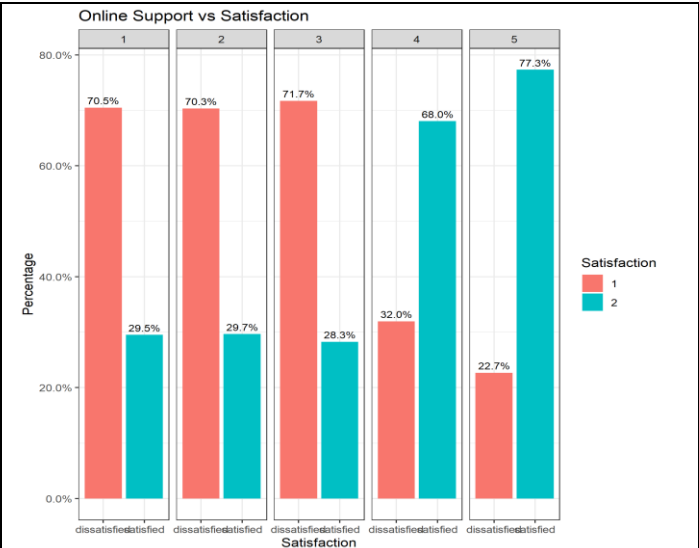
We observe that Online support has a significant effect on customer satisfaction since people with Rating 5 are 77.3% satisfied.


Baggage Handling vs Satisfaction

We observe that Baggage Handling has a significant effect on customer satisfaction since people with Rating 5 are 73.6% satisfied.

Inflight Entertainment vs Satisfaction

We observe that Inflight Entertainment has a significant effect on customer satisfaction since people with Rating 5 are 96.2% satisfied.

Correlation Plot

The correlation plot gives the correlation of the various parameters under study with each other in which different colours represent the intensity of correlation of one parameter with another; red showing negative correlation, white for no correlation and blue being used for showing positive correlation between the factors. The fadeness of the colour indicates the decrease in correlation.
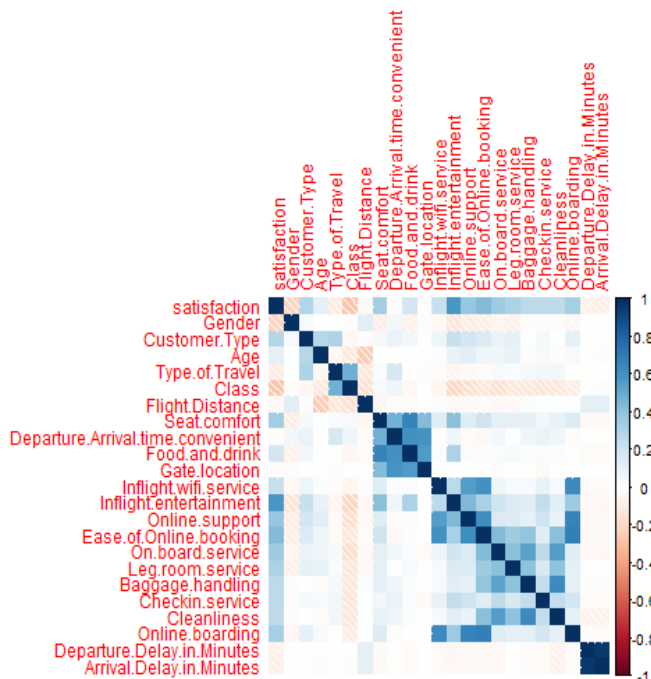
	Arrival.Delay.in.Minutes
satisfaction	-0.080690798
Gender	0.001308949
Customer.Type	-0.004730410
Age	-0.011247759
Type.of.Travel	-0.005829678
Class	0.014162417
Flight.Distance	0.110103111
Seat.comfort	-0.037875040
Departure.Arrival.time.convenient	-0.003367583
Food.and.drink	-0.022183992
Gate.location	0.003625964
Inflight.wifi.service	-0.028423814
Inflight.entertainment	-0.047152371
Online.support	-0.035954425
Ease.of.Online.booking	-0.039880950
On.board.service	-0.041509813
Leg.room.service	-0.001336796
Baggage.handling	-0.014164363
Checkin.service	-0.023756243
Cleanliness	-0.067273588
Online.boarding	-0.021666321
Departure.Delay.in.Minutes	0.965291184

We observe a very high correlation (0.965291184) between Departure.Delay.in.Minutes and Arrival.Delay.in.Minutes. Fortunately, without the need of replacing or omitting the NA values in the variable "Arrival.Delay.in.Minutes", we can exclude the variable entirely, as we have a similar variable having no NA values in our dataset to proceed with our modelling.

Logistic Regression Model
Call:
glm(formula = satisfaction ~ ., family = "binomial", data = train.lr)

Min	1Q	Median	3Q	Max					

					Estimate	Std.Error	z value	Pr(>|z|)
-3.2562	-0.4268	0.1533	0.4648	3.7217				
	Coefficients:							
					Estimate	Std.Error	z value	Pr(>|z|)
	(Intercept)				-8.590e+00	1.059e-01	-81.134	< 2e-16 ***
	Gender				-9.818e-01	2.154e-02	-45.571	< 2e-16 ***
	Customer.Type				2.211e+00	3.263e-02	67.755	< 2e-16 ***
	Age				-4.666e-03	7.467e-04	-6.249	4.13e-10 ***
	Type.of.Travel				-1.052e+00	2.871e-02	-36.636	< 2e-16 ***
	Class				-5.903e-01	2.001e-02	-29.493	< 2e-16 ***
	Flight.Distance				-8.957e-05	1.113e-05	-8.050	8.29e-16 ***
	Seat.comfort				6.410e-01	1.282e-02	49.989	< 2e-16 ***
	Departure.Arrival.time.convenient				-3.471e-01	1.126e-02	-30.834	< 2e-16 ***
	Food.and.drink				7.562e-02	1.356e-02	5.579	2.42e-08 ***
	Gate.location				-6.261e-02	1.077e-02	-5.812	6.17e-09 ***
	Inflight.wifi.service				-9.608e-02	1.153e-02	-8.334	< 2e-16 ***
	Inflight.entertainment				8.967e-01	1.166e-02	76.923	< 2e-16 ***
	Online.support				7.102e-02	1.184e-02	5.999	1.98e-09 ***
	Ease.of.Online.booking				2.085e-01	1.519e-02	13.722	< 2e-16 ***
	On.board.service				3.324e-01	1.092e-02	30.438	< 2e-16 ***
	Leg.room.service				2.533e-01	9.359e-03	27.069	< 2e-16 ***
	Baggage.handling				1.070e-01	1.217e-02	8.795	< 2e-16 ***
	Checkin.service				3.165e-01	9.051e-03	34.963	< 2e-16 ***
	Cleanliness				8.655e-02	1.271e-02	6.810	9.77e-12 ***
	Online.boarding				1.851e-01	1.283e-02	14.420	< 2e-16 ***
	Departure.Delay.in.Minutes				-4.480e-03	2.866e-04	-15.631	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1								
(Dispersion parameter for binomial family taken to be 1)								
Null deviance: 121980 on 88555 degrees of freedom								
Residual deviance: 59266 on 88534 degrees of freedom								
AIC: 59310								
Number of Fisher Scoring iterations: 6								
Goodness of fit test for p value:								
Hosmer and Lemeshow goodness of fit (GOF) test								
data: train.lr$satisfaction, fitted(logistic.model)								
X-squared = 88556, df = 8, p-value < 2.2e-16								

We observe that all independent variable are significantly important in predicting the "satisfaction". We predict the test dataset using the model built. The obtained predictions are in terms of probabilities and hence by considering a cutoff of 0.5(practically used default value), we classify the predicted "satisfaction" into "Yes" or "No". With the help of confusion matrix, the performance of the model is assessed using the predicted and observed values.

Confusion Matrix and Statistics

	Reference	
Prediction	**dissatisfied**	**satisfied**
• **dissatisfied**	15395	2770
• **satisfied**	3284	19875

Accuracy : 0.8535

From the above results, we can observe that:
1. The Overall accuracy of 85.35% and Kappa value of 70.36% indicates that our model performed better than a random prediction of the dependent variable.
2. The sensitivity value of 87.77% indicates that our model has correctly identified 87.77% of the satisfied customers correctly i.e the true positive rate

95% CI : (0.8501, 0.8569)
No Information Rate : 0.548
P-Value [Acc > NIR] : < 2.2e-16

**Kappa : 0.7036**

Mcnemar's Test P-Value : 4.305e-11

Sensitivity : 0.8777
Specificity : 0.8242
Pos Pred Value : 0.8582
Neg Pred Value : 0.8475
Precision : 0.8582
Recall : 0.8777
F1 : 0.8678
Prevalence : 0.5480
Detection Rate : 0.4810
Detection Prevalence : 0.5604
Balanced Accuracy : 0.8509

'Positive' Class : satisfied

3. The specificity value of 82.42% indicates that our model has identified 82.42% of the dissatisfied customers correctly i.e true negative rate

Decision Tree Algorithm
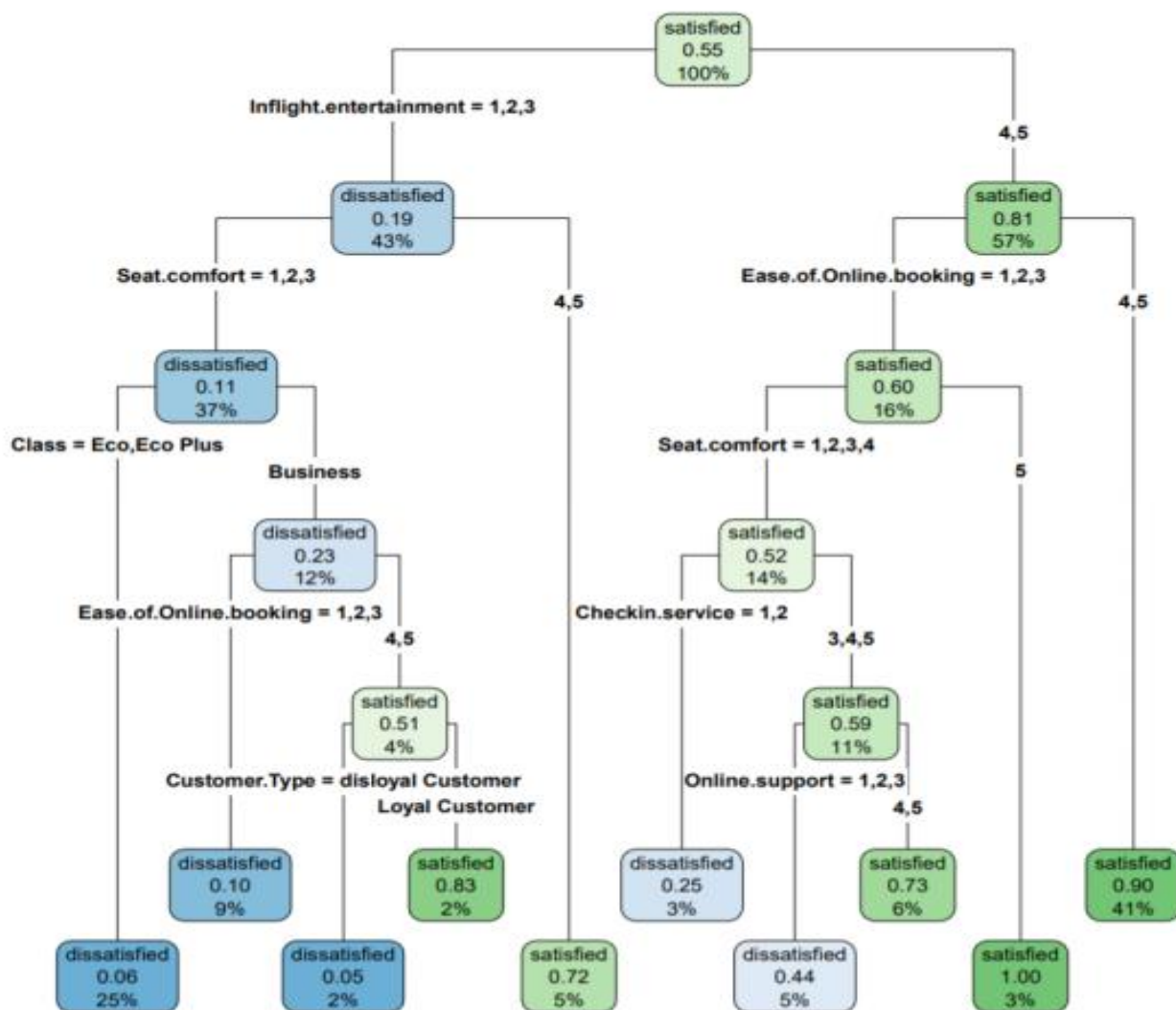
It is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. It uses a recursive partitioning method to divide the datasets into subsets repeatedly at each step.

We observe that the first or the more valuable variable we have "inflight entertainment" being rated 1-3 we classify them to be "dissatisfied" class on the left node, which is further classified based on the rating given on "Seat Comfort". If seat comfort is rated 1-3, it is further classified based on Class and Ease of Online Booking. If the Ease of Online Booking was is rated 1-3 and the class chosen was either Eco or Eco plus,we finally decided them to be dissatisfied. But if the Class was Business class, and Ease of Online Booking is rated 4-5, we classified based on Customer Type. If the customer is a disloyal customer, we decide them to be dissatisfied else satisfied. Similarly the tree progresses finally into 10 leaf nodes.

We use this model to predict the test data set and observe the confusion matrix to understand the model performance.

Confusion Matrix and Statistics

	Reference	
Prediction	**dissatisfied**	**satisfied**
• **dissatisfied**	**15374**	**2258**
• **satisfied**	**3305**	**20387**
•		

Accuracy : 0.8654
95% CI : (0.8621, 0.8687)
No Information Rate : 0.548
P-Value [Acc > NIR] : < 2.2e-16

From the above results, we can observe that:

1. The Overall accuracy of 86.54% and Kappa value of 72.69% indicates that our model performed better than a random prediction of the dependent variable.
2. The sensitivity value of 90.03% indicates that our model has correctly identified 90.03% of the satisfied customers correctly i.e the true positive rate
3. The specificity value of 82.31% indicates that our model has identified 82.31% of the dissatisfied customers correctly i.e true negative rate

Naive Bayes Model
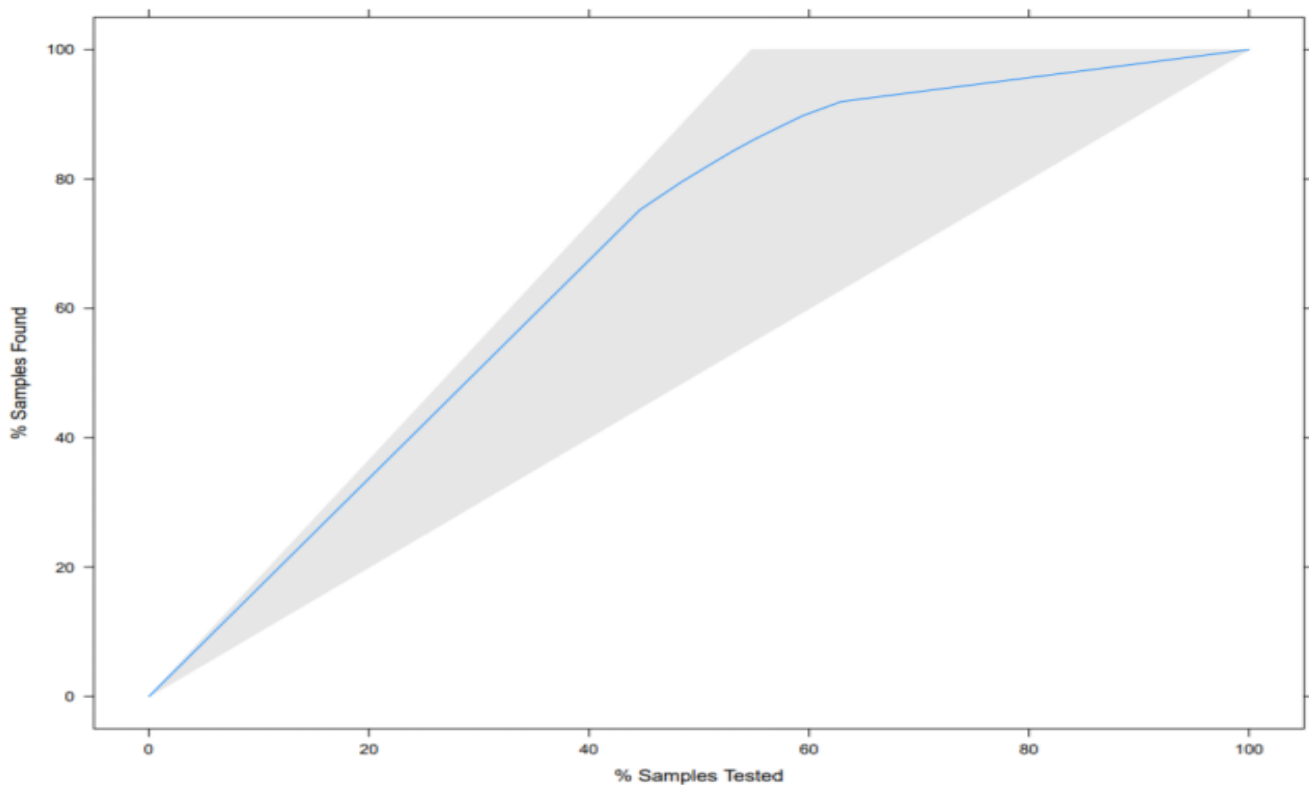
Bayesian Classifier - Lift Chart
The three lines denote the predicted, ideal and random line. Predicted line denotes the result of our model, Ideal line displays the ideal scenario whereas the random line displays results based on theoretical model which randomly select cases.

Confusion Matrix and Statistics		
Reference		
Prediction	**dissatisfied**	**satisfied**
• dissatisfied	15484	3323
• satisfied	3195	19322
Accuracy : 0.8423		

From the ove results we observe that:

1. The Overall accuracy of 84.23% and Kappa value of 68.18% indicates that our model performed better than a random prediction of the dependent variable.
2. The sensitivity value of 82.9% indicates that our model has correctly identified 82.9% of the

95% CI : (0.8387, 0.8458)
No Information Rate : 0.548
P-Value [Acc > NIR] : <2e-16

Kappa : 0.6818

Mcnemar's Test P-Value : 0.1157

Sensitivity : 0.8290
Specificity : 0.8533
Pos Pred Value : 0.8233
Neg Pred Value : 0.8581
Prevalence : 0.4520
Detection Rate : 0.3747
Detection Prevalence : 0.4551
Balanced Accuracy : 0.8411

'Positive' Class : dissatisfied

satisfied customers correctly i.e the true positive rate

3. The specificity value of 85.33% indicates that our model has identified 85.33% of the dissatisfied customers correctly i.e true negative rate

Limitations

Conclusion:

Model	Accuracy	Kappa	Sensitivity	Specificity
Logistic Regression	**85.35%**	**70.36%**	**87.77%**	**82.42%**
Decision Tree	86.54%	**72.69%**	90.03%	**82.31%**
Naïve Bayes Classification Algorithm	84.23%	**68.18%**	82.9%	**85.33%**

Through our analysis and referring to Table , we conclude that the decision tree built using the CART algorithm is statistically a better model for our dataset. This model provides us with an accuracy of 86.54% while it identifies 90.03% (sensitivity value) of the satisfied customers correctly i.e the true positive rate and 82.31%(specificity value) of the dissatisfied customers correctly i.e true negative rate.

We observe from the decision tree model, the Inflight Experience, Seat Comfort and Online Service levels significantly affect the customer experience and satisfaction level along with several other variables considered. The airline companies must ensure high quality service in these parameters to receive high customer satisfaction level. The models built lack to perfectly classify customer satisfaction level. The prediction accuracy can further be improved using advanced classification algorithms or deep learning concepts.

References

1. Kaggle dataset on Passenger Satisfaction for a US Airlines by John D. https://www.kaggle.com/johndddddd/customer-satisfaction
2. Usman Khan, Noreen Khan, Customer Satisfaction in Airline Industry (2014)
3. Dwi Suhartanto : The Department of Business Administration, Bandung State Polytechnic; Any Ariani Noor : The Department of Business Administration, Bandung State Polytechnic – Customer Satisfaction in the Airline Industry – The Role of Service Quality and Price (2012)
4. Zeithaml. V.A. Berry, L.L. & Parasuraman, The nature and determinants of customer expectations of service. Journal of the Academy of Marketing Science(1993)
5. Barsky, J., & Nash, L. Customer satisfaction: Applying concepts to industry-wide measures. Cornell Hotel and Restaurant Administration Quarterly, 44(5/6), 173.(2003)
6. Bhattacharya, S., & Singh, D.The emergence of hierarchy in customer perceived value for services: A grounded analysis. Journal of American Academy of Business, 13, 65.(2008)

7. Parasuraman. A., Berry. L.L,. & Zeithaml, V.A, Understanding customer expectations of service, Sloan Management Review, 32 (3), 39-48 (1991)
8. Hamer, L.O., Shaw-Ching Liu, B., & Sudharshan. D. The effects of intra encounter changes in expectations on perceived service quality models. Journal of Service Research, 1(3), 275-289.(1999)
9. Gronroos, C.Strategic management and marketing in the service sector. Cambridge. MA: Marketing Science Institute. International Journal of Service Industry Management 12(1): 5-10(1983)
10. Hill, D, J, Satisfaction and consumer services. In R.L.Lutz (Ed.). Advances in consumer research 13(1): 311-315. Ann Arbor, MI: Association for Consumer Research. (1986)
11. Richard. M.D, & Allaway, A.W. Service quality attributes and choice behavior. Journal of Services Marketing, 7(1), 59-68.(1993)
12. Walker, J. & Baker, J. An exploratory study of a multi-expectation framework for services. Journal of Services Marketing, 14(5), 411 - 431.(2000)
13. Strauss, B. & Schoeler, A. Complaint management profitability: what do complaint managers know? Managing Service Quality, 14(2):147-156. (2004)
14. Stauss, B. The dimensions of complaint satisfaction: process and outcome complaint satisfaction versus cold fact and warm act complaint satisfaction. Managing Service Quality, 12(3):173-183. (2002)
15. Palmatier, R. W. Integration of customer satisfaction and relationship marketing models for predicting repurchase intentions: an attribution approach. Proceeding of American Marketing Association Summer Educators' Conference. San Diego, CA, P. 163.(2002)
16. Davidow, M. Have you heard the word? The effect of word of mouth on perceived justice, satisfaction and repurchase intentions following complaint handling. Journal of Consumer Satisfaction, Dissatisfaction and Complaint Behavior, 16(1), 67-80.(2003)