INFO 430: Business Analytics

Airline Customer Satisfaction

Dominic Bartolomeo & Christopher Desmond

Introduction:

The following data analysis of the airlinequality.com customer satisfaction reports will provide a detailed review of customer opinions on various factors of several different airlines. The user-provided information such as the reviewer's name, how many reviews the reviewer made, the cabin class the reviewer flew (economy, business, or first class), what airline was used, and other reviewer-rated aspects of their flight. We can use this data to determine whether certain aspects of a reviewer's flight led to a "Yes" recommendation or a "No" recommendation. In the following analysis, we used several analytical data and text models to conclude our personal recommendations for the individual airlines to improve their customer satisfaction and retention based on our supported analytical analysis.

Data preparation:

The Airline Dataset contains quantitative data to represent individual customer satisfaction levels based on multiple factors. These factors are Value for money, Seat comfort, Catering, Entertainment, and whether or not the customers Recommend the airline. For our analysis, we decided to drop the 'URL' data as they are not significant to the analysis. - the URL data variable is not quantitative. We also noticed that there are multiple missing values within the data set that we plan to impute when running models.

The metadata is ready to be used based on the updated SAS studio data table. As seen in **Figure 2.1** we dropped 'URL' and re-coded Value for Money, Seat Comfort, Catering, Entertainment, and Cabin Flown into numeric data types, as they were character types before. In the process flow shown in **Figure 2.3,** we converted the provided Airline_Dataset.xlsx table to an Airline_Dataset.sas7bdat. In the process flow, we dropped the unnecessary data table including 'URLs'.

The target variable of the Airline dataset is the 'Recommended' data which identifies the overall opinion of the customer's level of satisfaction with the Airline. Other important variables include Value_for_money, Seat_comfort, Catering, and Entertainment. These variables are measured on a scale from 0-1, the higher the better. The distribution of the target data is different for every airline with some, AA and Emirate, skewing left, as seen in **Figures 2.10 & 2.12,** and having good ratings overall and a majority of their customers saying they would recommend the

airline. Others like Qantas skew right, as seen in **Figure 2.11**, and show low ratings across the board and most customers would not recommend these airlines. As for issues in the data we have noticed there are a few missing variables that must be converted inside the dataset.

3. **Data analysis:**

In **Figure 3.4**, we used Enterprise Miner to run four separate analysis models, we compared each model to find which was most significant. The four models we compared in our analysis were the Probability tree, Misclassification tree, Linear regression, and a Neural network. We used a model comparison node to evaluate the models based on the validation Average Square Error rate, which can be seen in **Figure 3.3**.

For our dataset, we used the "Recommended" variable as our target variable for the analysis. We select it as it shows whether or not the review from the customer was recommended as 'Yes' or 'No' and displays a great overall description of the review. To start our analysis we used a Data Partition node and a 70:30 split, 70 being for the training dataset and 30 being for the validation dataset, 0 for the test dataset. Then we added a Transform Variable node and rejected all nominal data. Following this we used the Impute node to reject the missing variables inside the dataset, we only connected the Impute node to the Regression and Neural Network nodes as they are sensitive to the missing values. We did not attach the impute node to the misclassification tree and the probability tree. To create our four models for analysis we created two decision tree nodes, one tree to identify the probability rate which has an assessment measure of Average Square Error, and another tree to measure the misclassification rate of the data assessment measure of decision. Connected to the impute node is the regression node which we modified to run a linear regression and a neural network, all of the models are connected to a model comparison node.

Based on the returned results for the model comparison node we determined that the Neural network model has the lowest ASE, of 0.055 in **Figure 3.3** & **Figure 3.2**, and was the selected model based on the model comparison ASE criteria. We found this information to be conclusive and recommended using the neural network model when analyzing the recommended target variable for the airline dataset as it has the lowest rate of average square error.

As shown in **Figure 3.5** there are 21 topics that produce from the Review variable. Some topics include the airlines, seats, and our user-rated variables, such as Entertainment and Cabin Flown information, as seen in **Figure 3.6**. In the process of finding the best model for predicting "Yes" we created the Flow Diagram shown in **Figure 3.10**. First, we used a Data Partition node to, as the name suggests, partition the data. We set the Training percentage to 75, Validation to 25, and Test to 0. Next for Text Parsing, we deleted the default table. For the Text Filter node, we changed the frequency weighting to Log and term weight to Inverse Document Frequency. For Metadata we changed all of the "text_topic_raw" data, as well as the reviewer data. For the models, Decision tree, Neural Network, and Regression, we changed the assessment method for the Decision Tree to ASE, then we changed Neural Network's model selection criterion to Average Error and changed the regression type to Linear regression for the Regression node. Finally running the model comparison, the model with the best ASE is the Decision Tree, as shown in **Figures 3.8 & 3.9**. In conclusion, the information in the review variable is useful to determine if a "Yes" will be the recommendation of any reviewer, specifically as shown in **Figure 3.10** the Seat Comfort is the most useful.

Results:

Based on our several analysis models of the airline dataset we have the following recommendations to make for airline company managers to improve customer satisfaction and retention rates. Using **Figure 3.11**, we can conclude the most discussed topic from all airline reviews is the level of seat comfort on the aircraft. Our analysis concludes that customers are inclined to be happier overall based on the level of seat comfort. The leaf nodes closely following the Seat Comfort is the catering rating, or level of customer satisfaction based on the airline-provided food, which seems to be a very high concern for customers across all airlines. This is also supported based on our analysis of **Figure 2.5** Qantas Summary Statistics, in the following summary statistic it can be identified that for the 'Recommended - No' division of the statistic Seat Comfort and Catering Rating are the highest levels of nominal data factors that affect the decision of customer satisfaction. The mean for the seat comfort value is .40 and the mean for the catering rating value is at .63 showing just how significant the following factors are to opinions of the customer recommendation based on the airline. This data is consistent with the other airline summary statistic figures shown in **Figures 2.4 - 2.9**.

We also identified the text topic of ‘cancellations, hour, delay’ as other important factors to consider as they are frequently discussed in the review text topics. The text topics ‘cancellations, hour, delay’ being important issues to customers is also supported inside **Figure 3.5 - Text Cluster Data**, as terms included inside the text topic, such as delay, occur many times inside the dataset. For example, cluster ID 20 includes terms such as the topics ‘delay, miss, late’ are mentioned with a frequency of 95, proving to be alarming to customers. Our findings report customers are very sensitive to time-sensitive obstacles as seen in the **Figure 3.11** decision tree when looking at the ‘< 0.7 or missing value’, reviews mentioning delays, cancellations, etc. are correlated with a low value for money and seem to make customers very frustrated.

Based on our analysis models we greatly urge airline managers to focus on Seat comfort and Catering quality (airline-provided food) to increase customer satisfaction and raise the customer opinions of the airline and the value for money factor of the airline. We also would recommend improving the operations of the airlines to prevent as many delays/cancellations as possible for the airline as customers are incredibly sensitive when time-sensitive topics occur.

Figures:

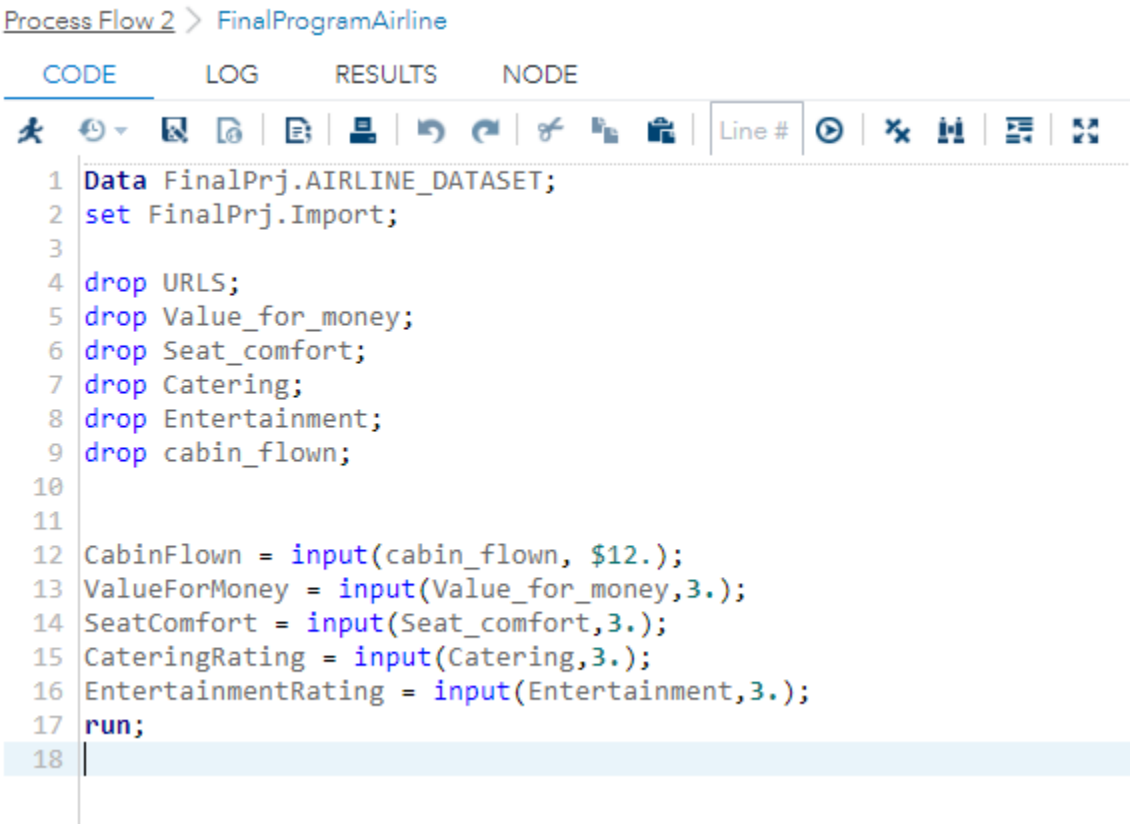

Figure 2.1 Recording Airline Data

OUTPUT DATA

Column names | Filter: (none)

Total rows: 1474 Total columns: 10 Rows 1-100

	Location	Recommen...	CabinFlown	ValueForMoney	SeatComfort	CateringRating
	USA	NO	Economy	0.4	0.2	0.6
	USA	YES	Premium Eco	1	1	0.8
	USA	YES	Economy	1	0.8	0.8
	USA	NO	Economy	0	0.2	0
	Panama	NO	Business	0.4	0.4	0.8
	USA	NO	Economy	0.2	0.6	0.2
	USA	NO	Economy	0.2	0.6	0.6
	USA	NO	First Class	0	0	0
	UK	YES	Premium Eco	1	0.8	0.6
	USA	NO	First Class	0.6	0.6	0.6
	USA	YES	Business	1	1	1
	USA	YES	Business	0.8	0.8	0.8
	New Zealand	NO	Economy	0.6	0.4	0.6
	USA	NO	Economy	0.2	0	0
	USA	YES	Premium Eco	0.8	0.8	0.6
	Canada	NO	Economy	0.4	0.2	0
	France	YES	Premium Eco	0.8	0.6	0.4
	USA	YES	Business	0.8	1	0.8
	USA	YES	Economy	0.8	0.8	0.8
	USA	NO	Economy	0.2	0.8	0
r	Netherlands	YES	Economy	0.8	0.6	0.6
	New Zealand	NO	Economy	0.6	0.8	0.4

Figure 2.2 Recoded Airline Data Chart

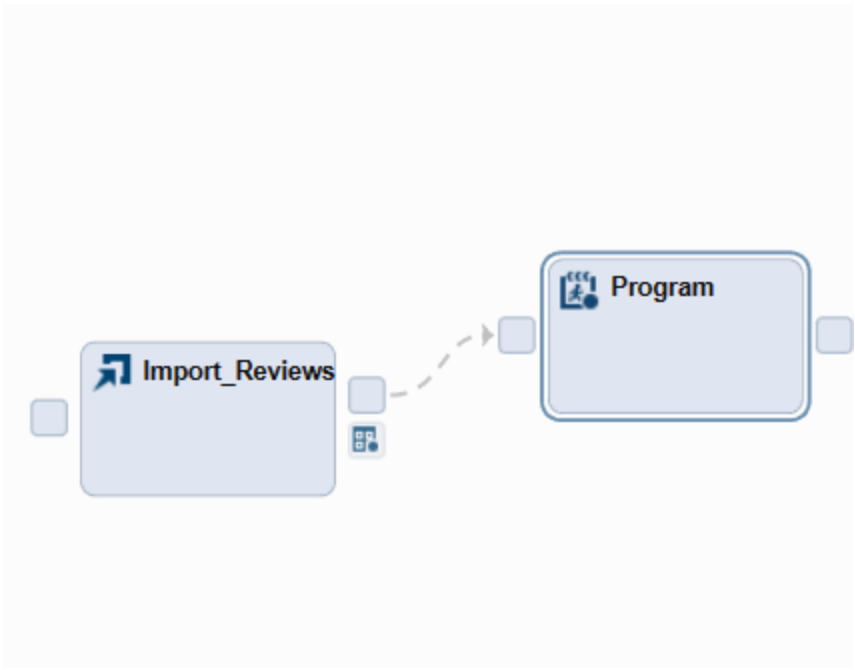

Figure 2.3 Reviews Process Flow

Airline=Emirates

Recommended	CabinFlown	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
NO	Business	15	ValueForMoney	0.4800000	0.1264911	0.2000000	0.6000000	15
			SeatComfort	0.6400000	0.2164651	0.2000000	1.0000000	15
			CateringRating	0.5333333	0.2225395	0.2000000	1.0000000	15
			EntertainmentRating	0.8000000	0.2390457	0.2000000	1.0000000	15
	Economy	67	ValueForMoney	0.4626866	0.1748036	0.2000000	0.8000000	67
			SeatComfort	0.4985075	0.2421448	0	1.0000000	67
			CateringRating	0.5104478	0.2182120	0.2000000	0.8000000	67
			EntertainmentRating	0.8507937	0.2093562	0.2000000	1.0000000	63
	First Class	3	ValueForMoney	0.4666667	0.3055050	0.2000000	0.8000000	3
			SeatComfort	0.6666667	0.4163332	0.2000000	1.0000000	3
			CateringRating	0.6000000	0.2000000	0.4000000	0.8000000	3
			EntertainmentRating	0.9333333	0.1154701	0.8000000	1.0000000	3
YES	Business	52	ValueForMoney	0.8423077	0.1741603	0	1.0000000	52
			SeatComfort	0.8269231	0.1483443	0.4000000	1.0000000	52
			CateringRating	0.8538462	0.1539818	0.4000000	1.0000000	52
			EntertainmentRating	0.8244898	0.2402521	0	1.0000000	49
	Economy	104	ValueForMoney	0.8384615	0.1626884	0.4000000	1.0000000	104
			SeatComfort	0.7500000	0.1757094	0.2000000	1.0000000	104
			CateringRating	0.8384615	0.1764305	0.2000000	1.0000000	104
			EntertainmentRating	0.8755102	0.1894437	0	1.0000000	98
	First Class	18	ValueForMoney	0.8888889	0.1231398	0.6000000	1.0000000	18
			SeatComfort	0.9777778	0.0646762	0.8000000	1.0000000	18
			CateringRating	0.8555556	0.2035630	0.4000000	1.0000000	18
			EntertainmentRating	0.7888889	0.3393573	0	1.0000000	18

Figure 2.4 Emirates Summary Statistics

Airline=Qantas

Recommended	CabinFlown	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
NO	Business	14	ValueForMoney	0.4000000	0.2075498	0.2000000	0.8000000	14
			SeatComfort	0.4428571	0.2502746	0.2000000	1.0000000	14
			CateringRating	0.5428571	0.2533295	0	1.0000000	14
			EntertainmentRating	0.6428571	0.3694368	0	1.0000000	14
	Economy	26	ValueForMoney	0.4307692	0.2168303	0	1.0000000	26
			SeatComfort	0.4307692	0.1934306	0.2000000	0.8000000	26
			CateringRating	0.6000000	0.2939388	0.2000000	1.0000000	26
			EntertainmentRating	0.7769231	0.2486889	0	1.0000000	26
	First Class	2	ValueForMoney	0.4000000	0.2828427	0.2000000	0.6000000	2
			SeatComfort	0.7000000	0.4242641	0.4000000	1.0000000	2
			CateringRating	0.7000000	0.1414214	0.6000000	0.8000000	2
			EntertainmentRating	0.6000000	0.2828427	0.4000000	0.8000000	2
	Premium Eco	2	ValueForMoney	0.3000000	0.1414214	0.2000000	0.4000000	2
			SeatComfort	0.4000000	0	0.4000000	0.4000000	2
			CateringRating	0.6000000	0	0.6000000	0.6000000	2
			EntertainmentRating	0.8000000	0.2828427	0.6000000	1.0000000	2
YES	Business	56	ValueForMoney	0.8464286	0.1788491	0.2000000	1.0000000	56
			SeatComfort	0.8714286	0.1723783	0.2000000	1.0000000	56
			CateringRating	0.8145455	0.2067546	0.2000000	1.0000000	55
			EntertainmentRating	0.7773585	0.2736094	0	1.0000000	53
	Economy	99	ValueForMoney	0.8686869	0.1375043	0.4000000	1.0000000	99
			SeatComfort	0.7898990	0.1837783	0.2000000	1.0000000	99
			CateringRating	0.7717172	0.2490965	0	1.0000000	99
			EntertainmentRating	0.7814433	0.2785116	0	1.0000000	97
	First Class	4	ValueForMoney	0.8500000	0.1000000	0.8000000	1.0000000	4
			SeatComfort	1.0000000	0	1.0000000	1.0000000	4
			CateringRating	1.0000000	0	1.0000000	1.0000000	4
			EntertainmentRating	0.8666667	0.1154701	0.8000000	1.0000000	3
	Premium Eco	20	ValueForMoney	0.8500000	0.1701393	0.4000000	1.0000000	20
			SeatComfort	0.8300000	0.1490320	0.4000000	1.0000000	20
			CateringRating	0.8600000	0.2257152	0	1.0000000	20
			EntertainmentRating	0.8555556	0.1503808	0.6000000	1.0000000	18

Figure 2.5 Qantas Summary Statistics

Airline=Singapore Airline

Recommended	CabinFlown	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
NO	Business	6	ValueForMoney	0.4000000	0.1788854	0.2000000	0.6000000	6
			SeatComfort	0.4000000	0.2190890	0.2000000	0.6000000	6
			CateringRating	0.6333333	0.4273952	0	1.0000000	6
			EntertainmentRating	0.8333333	0.3204164	0.2000000	1.0000000	6
	Economy	32	ValueForMoney	0.4687500	0.2070336	0	0.8000000	32
			SeatComfort	0.5187500	0.2428826	0	0.8000000	32
			CateringRating	0.7000000	0.2962127	0	1.0000000	32
			EntertainmentRating	0.7103448	0.3277029	0	1.0000000	29
	First Class	1	ValueForMoney	0.4000000	.	0.4000000	0.4000000	1
			SeatComfort	1.0000000	.	1.0000000	1.0000000	1
			CateringRating	0	.	0	0	1
			EntertainmentRating	1.0000000	.	1.0000000	1.0000000	1
YES	Business	33	ValueForMoney	0.8181818	0.2256304	0.2000000	1.0000000	33
			SeatComfort	0.8666667	0.1779513	0.2000000	1.0000000	33
			CateringRating	0.8258065	0.1611785	0.4000000	1.0000000	31
			EntertainmentRating	0.8437500	0.1664380	0.4000000	1.0000000	32
	Economy	103	ValueForMoney	0.8679612	0.1415693	0.4000000	1.0000000	103
			SeatComfort	0.8155340	0.1829804	0	1.0000000	103
			CateringRating	0.8333333	0.1991732	0.2000000	1.0000000	102
			EntertainmentRating	0.8257426	0.2377618	0	1.0000000	101
	First Class	4	ValueForMoney	0.8500000	0.1914854	0.6000000	1.0000000	4
			SeatComfort	0.9000000	0.2000000	0.6000000	1.0000000	4
			CateringRating	0.9000000	0.1154701	0.8000000	1.0000000	4
			EntertainmentRating	0.4500000	0.3415650	0	0.8000000	4
	Premium Eco	1	ValueForMoney	0.6000000	.	0.6000000	0.6000000	1
			SeatComfort	0.6000000	.	0.6000000	0.6000000	1
			CateringRating	0.8000000	.	0.8000000	0.8000000	1
			EntertainmentRating	1.0000000	.	1.0000000	1.0000000	1

Figure 2.6 Singapore Airlines Summary Statistics

Airline=Southwest

Recommended	CabinFlown	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
NO	Business	1	ValueForMoney	0.8000000	.	0.8000000	0.8000000	1
			SeatComfort	0.6000000	.	0.6000000	0.6000000	1
			CateringRating	0.4000000	.	0.4000000	0.4000000	1
			EntertainmentRating	.	.	.	.	0
	Economy	35	ValueForMoney	0.3485714	0.1960964	0.2000000	0.8000000	35
			SeatComfort	0.4457143	0.1945475	0	0.6000000	35
			CateringRating	0.3657143	0.2350755	0	0.8000000	35
			EntertainmentRating	.	.	.	.	0
YES	Cabin	1	ValueForMoney	1.0000000	.	1.0000000	1.0000000	1
			SeatComfort	1.0000000	.	1.0000000	1.0000000	1
			CateringRating	0.8000000	.	0.8000000	0.8000000	1
			EntertainmentRating	.	.	.	.	0
	Economy	104	ValueForMoney	0.9230769	0.1679290	0	1.0000000	104
			SeatComfort	0.8230769	0.1875912	0	1.0000000	104
			CateringRating	0.6576923	0.3432264	0	1.0000000	104
			EntertainmentRating	.	.	.	.	0

Figure 2.7 Southwest Summary Statistics

Airline=US Airways

Recommended	CabinFlown	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
NO	Business	5	ValueForMoney	0.3600000	0.1673320	0.2000000	0.6000000	5
			SeatComfort	0.4800000	0.2280351	0.2000000	0.8000000	5
			CateringRating	0.2000000	0.1414214	0	0.4000000	5
			EntertainmentRating	0.4000000	0.3162278	0	0.8000000	5
	Economy	87	ValueForMoney	0.3149425	0.1895703	0	0.8000000	87
			SeatComfort	0.4252874	0.2511455	0	1.0000000	87
			CateringRating	0.3057471	0.2195960	0	0.8000000	87
			EntertainmentRating	0.2942529	0.3032216	0	1.0000000	87
	First Class	8	ValueForMoney	0.1750000	0.1281740	0	0.4000000	8
			SeatComfort	0.4750000	0.2815772	0.2000000	1.0000000	8
			CateringRating	0.2000000	0.2138090	0	0.6000000	8
			EntertainmentRating	0.2000000	0.1851640	0	0.6000000	8
YES	Business	11	ValueForMoney	0.7818182	0.1401298	0.6000000	1.0000000	11
			SeatComfort	0.8727273	0.1009050	0.8000000	1.0000000	11
			CateringRating	0.8363636	0.1206045	0.6000000	1.0000000	11
			EntertainmentRating	0.4800000	0.3011091	0	0.8000000	10
	Economy	62	ValueForMoney	0.8548387	0.1210303	0.6000000	1.0000000	62
			SeatComfort	0.7580645	0.1408969	0.4000000	1.0000000	62
			CateringRating	0.6032258	0.2673291	0	1.0000000	62
			EntertainmentRating	0.3666667	0.4048296	0	1.0000000	60
	First Class	10	ValueForMoney	0.8200000	0.1135292	0.6000000	1.0000000	10
			SeatComfort	0.7600000	0.1264911	0.6000000	1.0000000	10
			CateringRating	0.8200000	0.1475730	0.6000000	1.0000000	10
			EntertainmentRating	0.2800000	0.3293090	0	0.8000000	10

Figure 2.8 US Airways Summary Statistics

Airline=United Airlines

Recommended	CabinFlown	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	N
NO	Business	15	ValueForMoney	0.3733333	0.2120198	0	0.8000000	15
			SeatComfort	0.5733333	0.2374467	0.2000000	1.0000000	15
			CateringRating	0.4285714	0.3023716	0	1.0000000	14
			EntertainmentRating	0.5866667	0.3583029	0	1.0000000	15
	Economy	133	ValueForMoney	0.3774436	0.1901421	0	0.8000000	133
			SeatComfort	0.4481203	0.2311373	0	0.8000000	133
			CateringRating	0.4150376	0.2802856	0	1.0000000	133
			EntertainmentRating	0.4854962	0.3356102	0	1.0000000	131
	First Class	7	ValueForMoney	0.3428571	0.2225395	0	0.6000000	7
			SeatComfort	0.4857143	0.3804759	0	1.0000000	7
			CateringRating	0.2571429	0.2507133	0	0.6000000	7
			EntertainmentRating	0.3142857	0.2544836	0	0.6000000	7
	Premium Eco	27	ValueForMoney	0.3629630	0.2150995	0	0.8000000	27
			SeatComfort	0.4074074	0.2384661	0	0.8000000	27
			CateringRating	0.3703704	0.2812265	0	1.0000000	27
			EntertainmentRating	0.3851852	0.3634278	0	1.0000000	27
YES	Business	24	ValueForMoney	0.8250000	0.1359348	0.6000000	1.0000000	24
			SeatComfort	0.8500000	0.1351328	0.4000000	1.0000000	24
			CateringRating	0.6333333	0.3212295	0	1.0000000	24
			EntertainmentRating	0.6083333	0.2796219	0.2000000	1.0000000	24
	Economy	40	ValueForMoney	0.8300000	0.1399634	0.6000000	1.0000000	40
			SeatComfort	0.7200000	0.1488417	0.4000000	1.0000000	40
			CateringRating	0.6200000	0.2633122	0	1.0000000	40
			EntertainmentRating	0.4205128	0.3172078	0	1.0000000	39
	First Class	11	ValueForMoney	0.8181818	0.1401298	0.6000000	1.0000000	11
			SeatComfort	0.8545455	0.0934199	0.8000000	1.0000000	11
			CateringRating	0.6545455	0.2696799	0	1.0000000	11
			EntertainmentRating	0.2727273	0.3495452	0	1.0000000	11
	Premium Eco	26	ValueForMoney	0.8461538	0.1303250	0.6000000	1.0000000	26
			SeatComfort	0.8230769	0.1305609	0.6000000	1.0000000	26
			CateringRating	0.5384615	0.2246365	0	1.0000000	26
			EntertainmentRating	0.3769231	0.3409489	0	1.0000000	26

Figure 2.9 United Airlines Summary Statistics

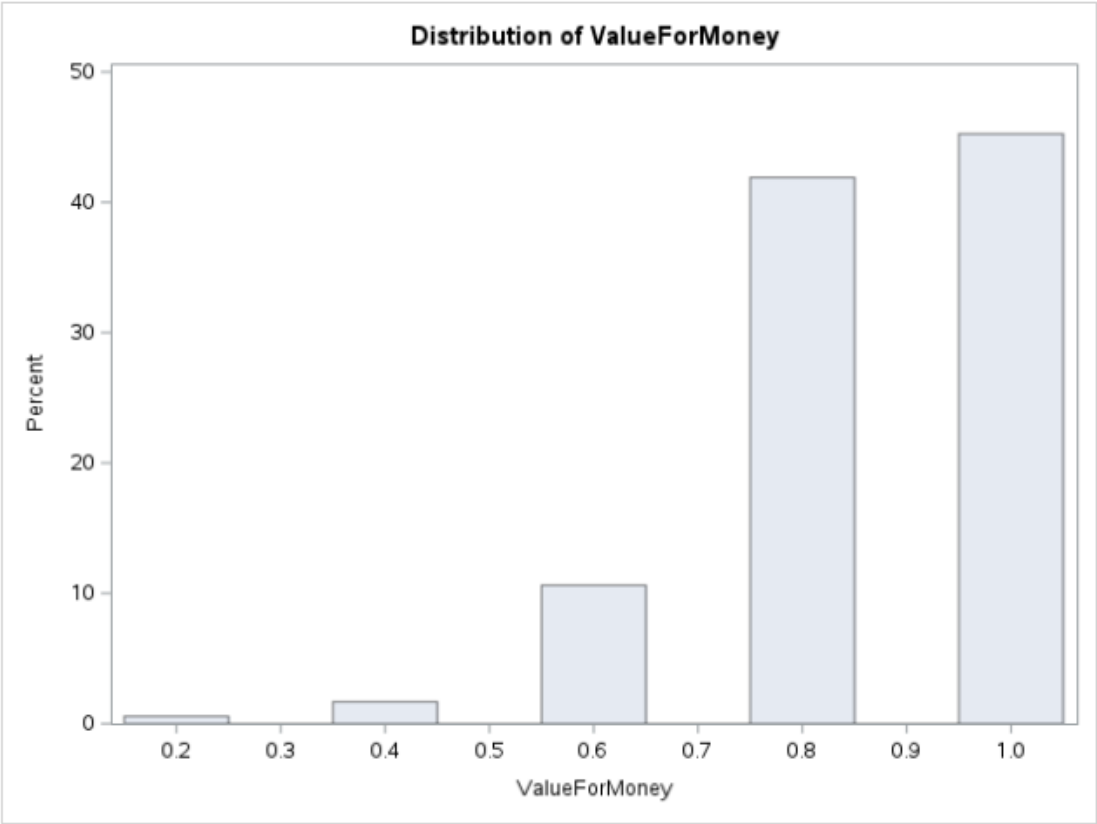

Figure 2.10 American Airline Recommended = YES

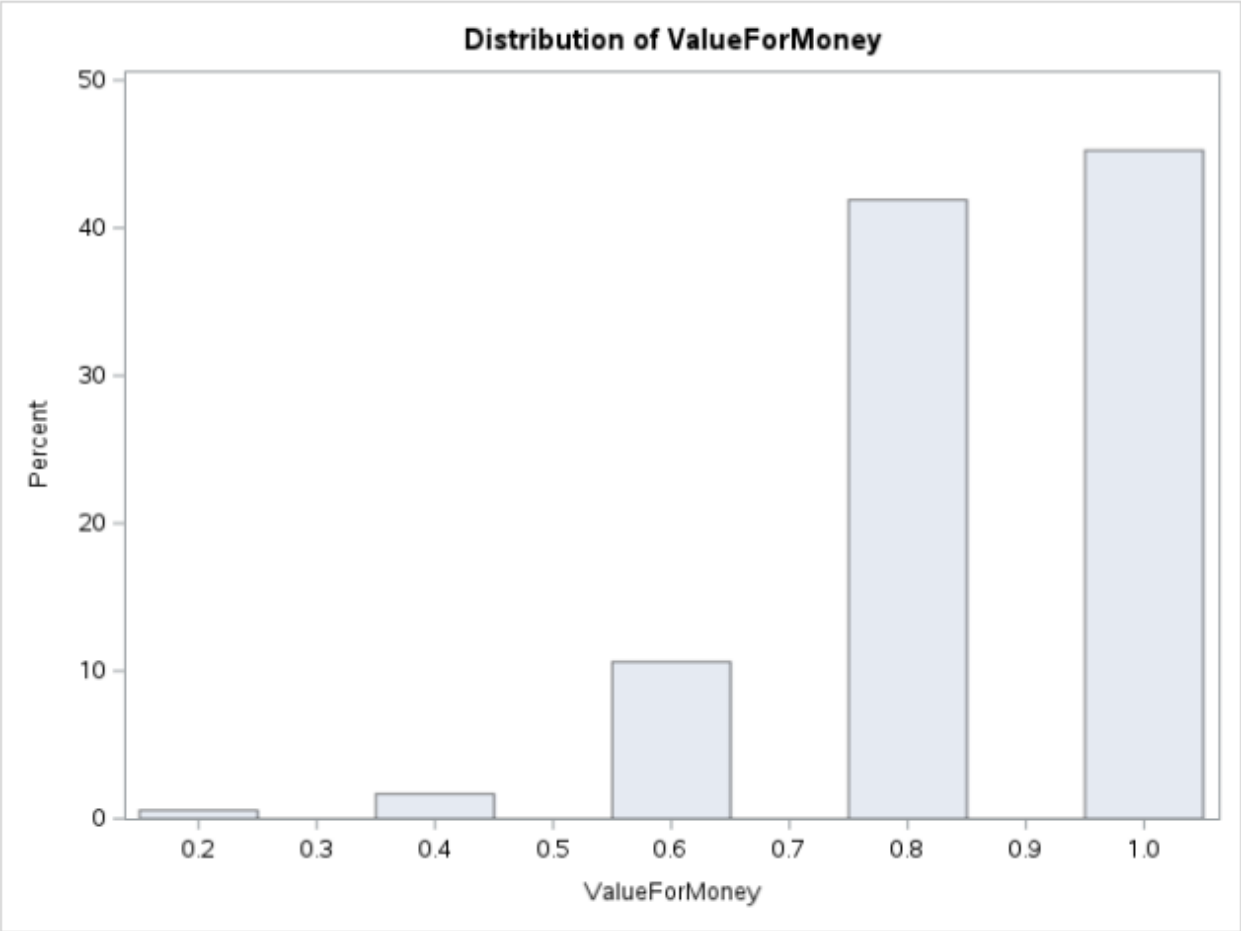

Figure 2.11 Qantas Airline Recommended = YES

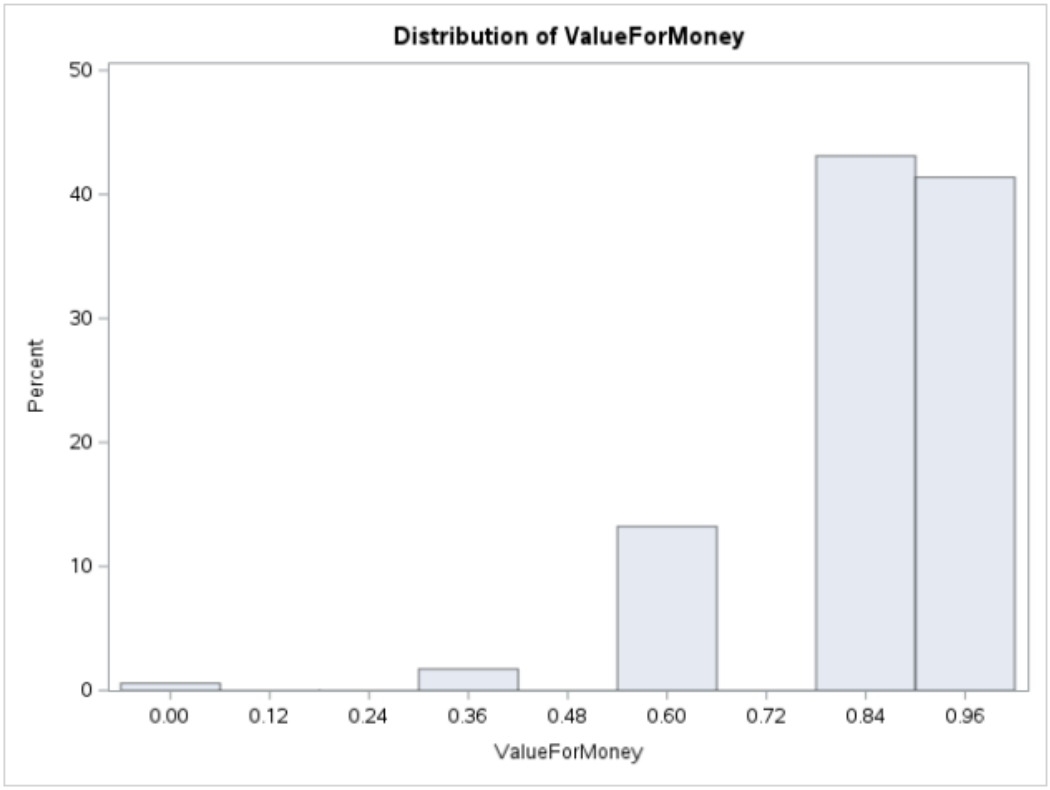

Figure 2.12 Emirates Airline Recommended = YES

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
Recommended	Recommended	_DFT_	Total Degrees of Free...	1031	.
Recommended	Recommended	_DFE_	Degrees of Freedom f...	1006	.
Recommended	Recommended	_DFM_	Model Degrees of Fre...	25	.
Recommended	Recommended	_NW_	Number of Estimated ...	25	.
Recommended	Recommended	_AIC_	Akaike's Information C...	415.4518	.
Recommended	Recommended	_SBC_	Schwarz's Bayesian C...	538.9089	.
Recommended	Recommended	_ASE_	Average Squared Error	0.05256	0.056045
Recommended	Recommended	_MAX_	Maximum Absolute Err...	0.99444	0.985544
Recommended	Recommended	_DIV_	Divisor for ASE	2062	886
Recommended	Recommended	_NOBS_	Sum of Frequencies	1031	443
Recommended	Recommended	_RASE_	Root Average Squared...	0.229259	0.236738
Recommended	Recommended	_SSE_	Sum of Squared Errors	108.3779	49.65597
Recommended	Recommended	_SUMW_	Sum of Case Weights ...	2062	886
Recommended	Recommended	_FPE_	Final Prediction Error	0.055172	.
Recommended	Recommended	_MSE_	Mean Squared Error	0.053866	0.056045
Recommended	Recommended	_RFPE_	Root Final Prediction ...	0.234887	.
Recommended	Recommended	_RMSE_	Root Mean Squared E...	0.23209	0.236738
Recommended	Recommended	_AVERR_	Average Error Function	0.177232	0.188452
Recommended	Recommended	_ERR_	Error Function	365.4518	166.9683
Recommended	Recommended	_MISC_	Misclassification Rate	0.074685	0.074492
Recommended	Recommended	_WRONG_	Number of Wrong Cla...	77	33

Figure 3.1 Fit Statistics for Model Comparison w/o Text Mining

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	Train: Sum of Squared Errors	Train: Average Squared Error
Y	Neural	Neural	Neural Net...	Recommen...	Recommen...	0.074492	1031	0.074685	0.99444	108.3779	0.05256
	Tree	Tree	probability t...	Recommen...	Recommen...	0.079007	1031	0.077595	0.996063	119.2823	0.057848
	Tree2	Tree2	misclassifi...	Recommen...	Recommen...	0.079007	1031	0.077595	0.971519	125.6743	0.060948
	Reg	Reg	Regression	Recommen...	Recommen...	0.079007	1031	0.080504	1.452137	80.1225	0.077713

Figure 3.2 Fit Statistics Showing the Best Comparison

Valid: Average Squared Error
0.05576
0.05982
0.060817
0.06771

Figure 3.3 Model Comparison ASE w/o Text Mining

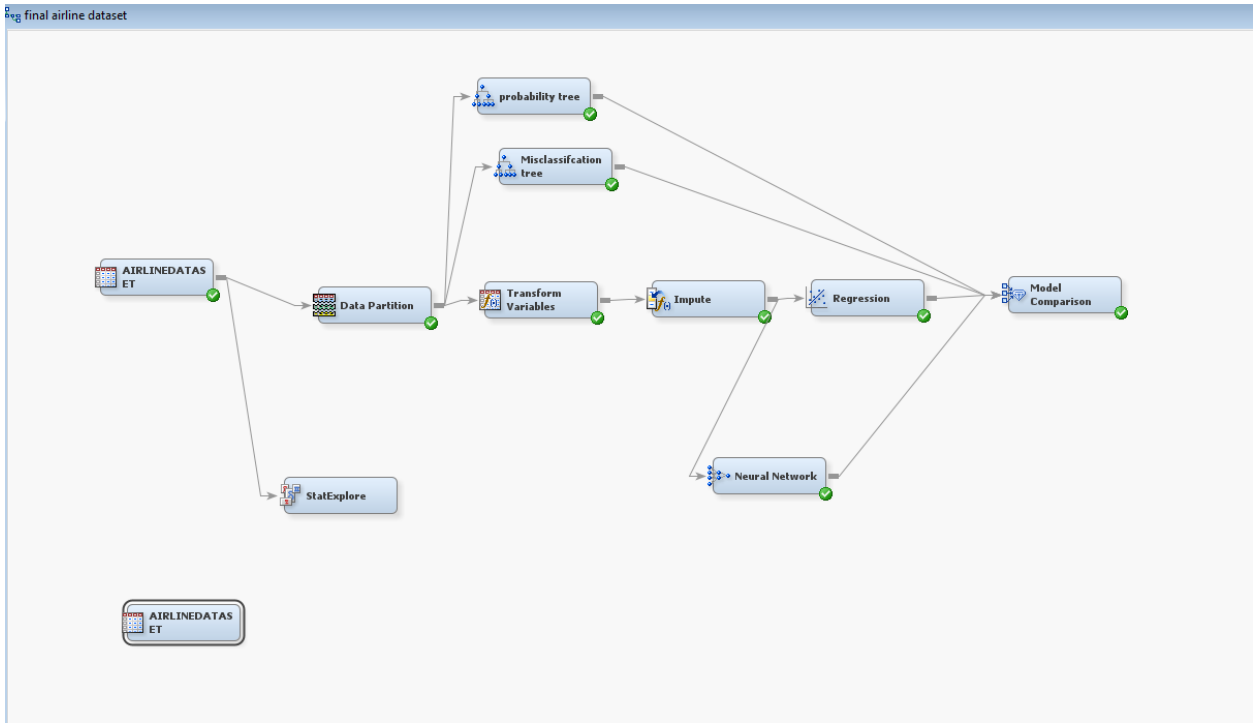

Figure 3.4 Flow Diagram w/o Text Mining

Cluster ID	Descriptive Terms	Frequency	Percentage
1	swa southwest open +know +reason +agree +policy +fee +cheap +choose +fare free luck...	11	1%
2	+delay +cancel +day +connect +miss +'connecting flight' +hotel next +book +hour morning ...	114	8%
3	'on time' +early +arrive overall +'all flight' +flight clean +friendly great return +crew +snack ...	114	8%
4	earlybird +access 24hours +chip +check-in +month +book advance 'at all' +mile +happen ...	12	1%
5	guess vegas southwest simply lucky +kind +hear +employee +plane american uncomforta...	16	1%
6	a380 excellent +lounge dubai dxb +crew ife always +food +drink 'a bit' comfortable emirate...	119	8%
7	united weather chicago +pass san +employee +customer luggage diego lax vegas +check...	84	6%
8	dxb +transfer +terminal +poor dubai +connection emirates +late +desk +minute +long +wa...	36	2%
9	singapore +great melbourne a380 sydney airlines attentive qantas excellent +service +frien...	110	7%
10	+child +sit son +young +old +daughter +year +family +age +attendant southwest +move +...	36	2%
11	+leg +room 'leg room' +aisle +seat +pay +drink +attendant +fee +upgrade first clean +frie...	138	9%
12	airways charlotte +side clean +easy +review airlines +'all flight' +ticket +airline 'on time' +fe...	24	2%
13	+bag checked 'carry on' free +check luggage +want baggage +snack +group check south...	65	4%
14	+class business 'business class' first +lounge +bed +choice flat +upgrade class +seat +p...	107	7%
15	class +bed business flat first comfortable +lounge attentive +access +wine good aircraft +...	50	3%
16	entertainment emirates +'entertainment system' inflight dubai +system +food +meal +feel +t...	138	9%
17	aa jfk +tv +bad +kind +movie +experience +hear +old +transfer 'at all' +fly +food +poor lu...	60	4%
18	+long +haul +'long haul' pleasant 'good service' international +cheap +fare +review +airline ...	57	4%
19	economy qantas premium 'upper deck' upper deck +wine +feel sydney +serve ife +cabin ...	59	4%
20	+delay +gate +wait finally +connection +minute +late +sit +hour +plane +agent +arrive +d...	95	6%
21	san +'gate agent' funny jose ca boarding +agent diego +gate +area +group free +desk sw...	29	2%

Figure 3.5 Text Cluster Data

Topic	Category	Term Cutoff	Document Cutoff	Number of Terms	# Docs
on time,+arrive,+pay,+flight,return	Multiple	0.031	0.115	167	115
+cancel,+hour,+delay,+day,+flight	Multiple	0.03	0.117	137	145
dubai,emirates,+airport,dxb,ice	Multiple	0.03	0.092	146	130
business,+class,business class,+seat,+lounge	Multiple	0.03	0.086	126	116
staff,+seat,+leg,+check-in,+room	Multiple	0.031	0.072	166	112
+attendant,+drink,inflight,+flight attendant,boarding	Multiple	0.031	0.07	191	66
+system,entertainment,united,+entertainment system,on time	Multiple	0.031	0.093	140	164
sydney,+great,qantas,a380,singapore	Multiple	0.03	0.127	115	158
return,boarding,return,+pass,+hour	Multiple	0.031	0.088	156	122
+cabin,+crew,+cabin crew,+travel,+crew	Multiple	0.031	0.08	145	140
staff,+good,+long,+delay,pleasant	Multiple	0.031	0.099	157	121
+meal,+serve,dinner,+attendant,breakfast	Multiple	0.031	0.085	189	157

Figure 3.6 Text Topic Data

Name	Role	Level	Report	Order	Drop
Airline	Input	Nominal	No		No
CabinFlown	Input	Nominal	No		No
CateringRating	Input	Interval	No		No
Date	Time ID	Interval	No		No
EntertainmentRa	Input	Interval	No		No
Location	Input	Nominal	No		No
Recommended	Input	Nominal	No		No
Review	Text	Nominal	No		No
Reviewer	Input	Nominal	No		No
SeatComfort	Input	Interval	No		No
ValueForMoney	Input	Interval	No		No

Figure 3.7 Metadata Information

Valid: Average Squared Error
0.102341
0.112089
0.121607

Figure 3.8 Validation ASE for Text Mining Model Comparison

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label
Y	Tree	Tree	Decision ...	Recomm...	Recomm...
	Neural	Neural	Neural N...	Recomm...	Recomm...
	Reg	Reg	Regressi...	Recomm...	Recomm...

Figure 3.9 Model Comparison for Text Mining

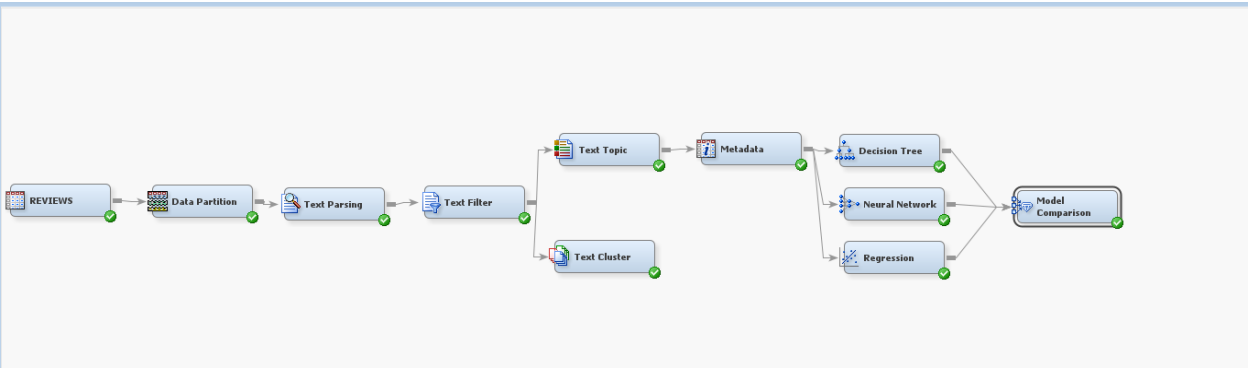

Figure 3.10 Process Flow for Text Mining Analysis

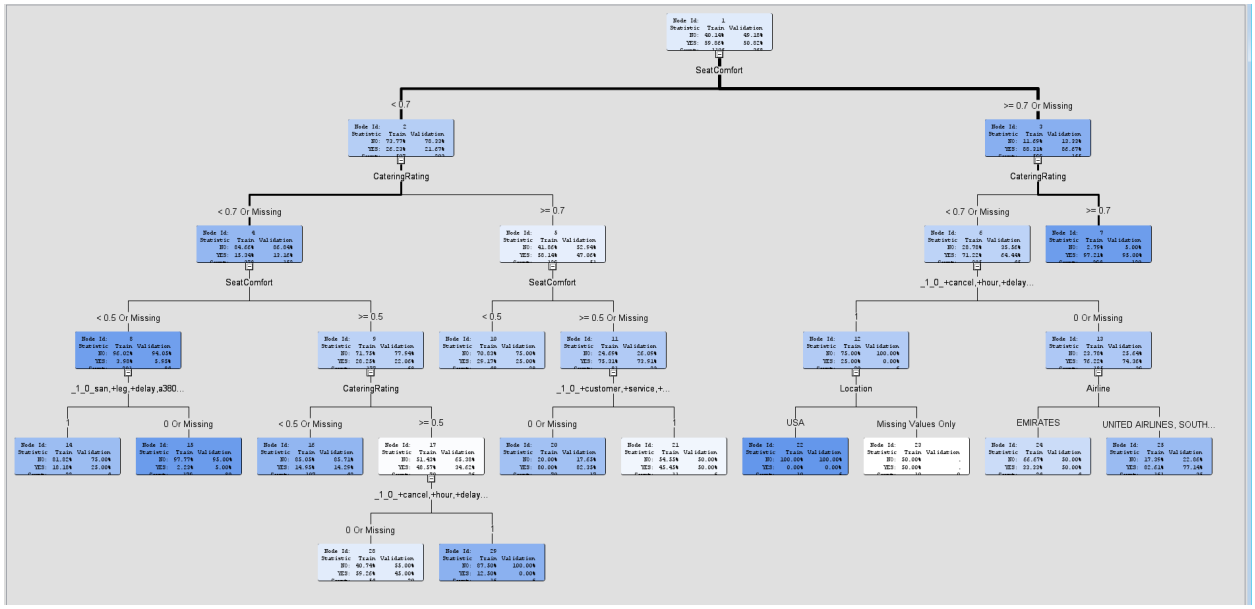

Figure 3.11 Text Mining Decision Tree