

AIRLINE

AIRLINE OPERATIONAL IMPROVEMENT RECOMMENDATION AND CUSTOMERS EXPERIENCE REVIEWS ANALYSIS

Group Name: BD Expert

Name	SID
Tsao Lyu Han	56119560
NG CHING WAI	56459370
CHONG HO HIN	56893610
Yan Man Ho Chris	56751318

TABLE OF CONTENTS

01

INTRODUCTION

SKYTRAX Introduction

04

OBJECTIVE 3

identify the critical factor that affect customer's recommendation

02

OBJECTIVE 1

To identify the topics of the positive and negative reviews

05

OBJECTIVE 4

Identify the aircraft components that consistently experience deflection

03

OBJECTIVE 2

To analyze how customers feel about airline service in a more detailed way

06

APPENDIX





01

INTRODUCTION

Company Introduction

COMPANY INTRODUCTION

The Company:

- Consulting firm that runs an airline and airport review and ranking site
- An airline forum where passengers give potential passengers insights and opinions about an airline

Research Purposes:

- Identify points travellers care the most
- Service evaluation





02

OBJECTIVE 1

To identify the topics of the positive and negative reviews

OBJECTIVE DETAILS

Objective

To investigate text from positive and negative reviews to understand primary topics within those comments

Methodology

Clustering
Topic Extraction

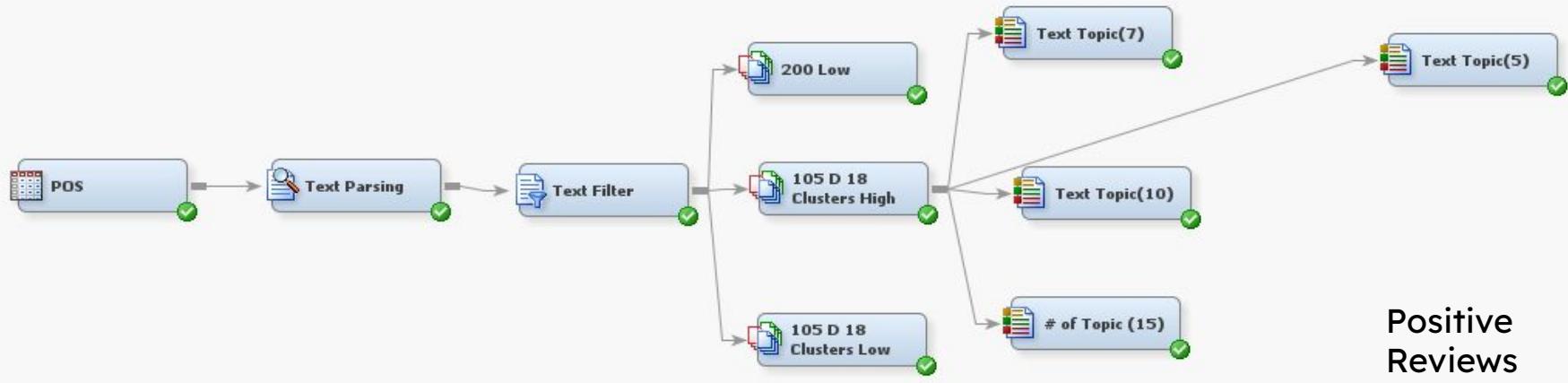
Data Input

Positive_Review
Negative_Review

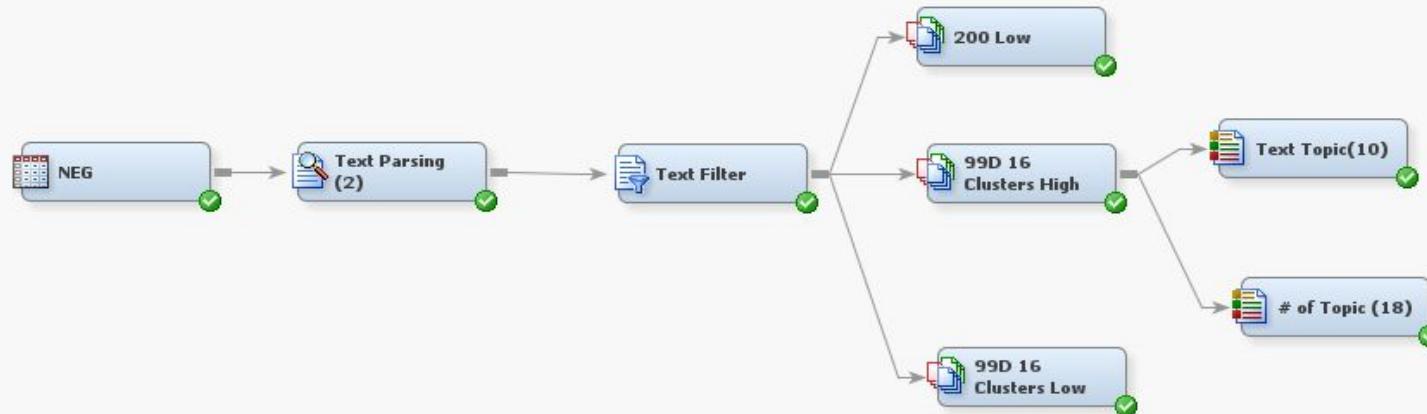
DATA PREPARATION

- To increase the accuracy of analysis
 - Use the Excel filter function to filter out some no meaningful and useless reviews
 - For example: no positive, no negative, wrong word, nope, none and so on

DIAGRAM OVERVIEW

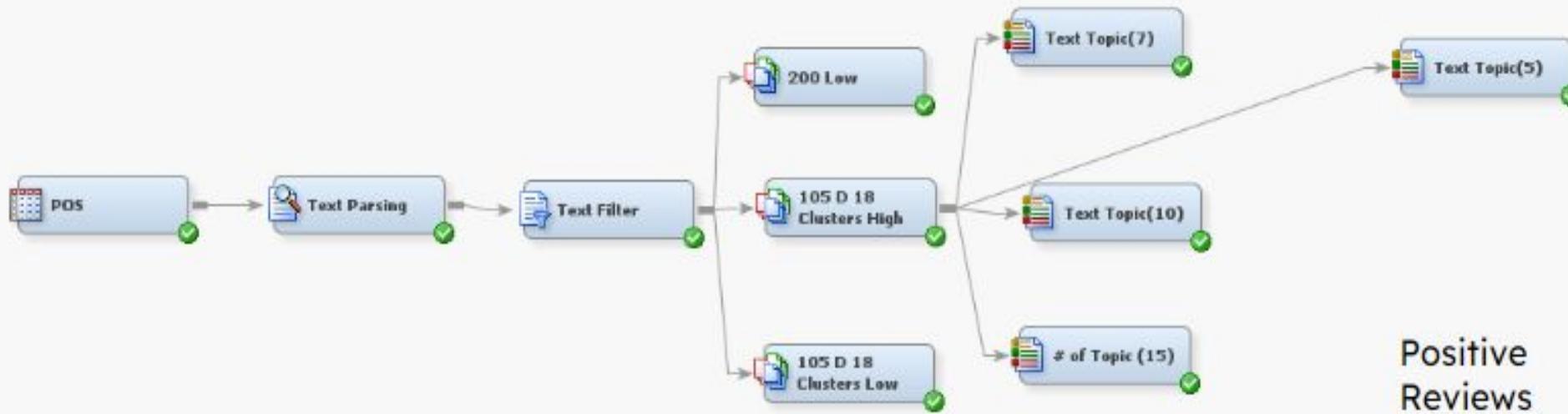


Positive
Reviews

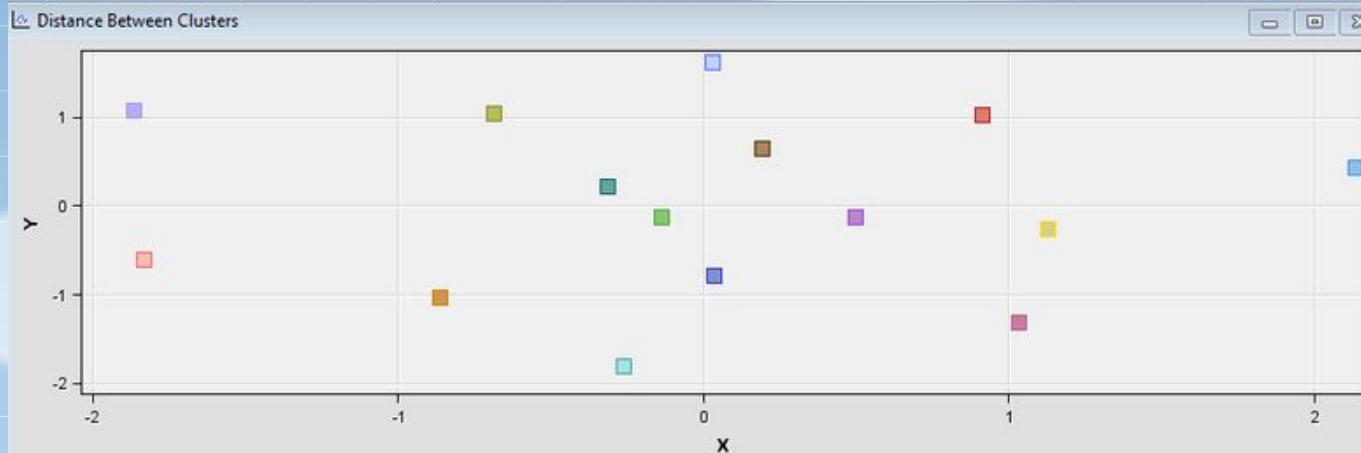


Negative
Reviews

POSITIVE REVIEWS



POSITIVE REVIEWS TEXT CLUSTERS



Cluster ID	Descriptive Terms	Frequency	Percentage
1	+fly +price +pay +ticket airways +book +review +cheap always +recommend +trip +travel +food +service +luggage	620	6%
2	+long +haul return +'long haul' +'return flight' +short +'haul flight' +flight economy +recommend great +leg airways +good +sy...	469	5%
3	+lounge 'on time' +check boarding +arrive +minute +airport +board +bed +wait full +luggage first +time +day	1821	18%
4	+serve +drink +breakfast +hot +dinner +snack +landing +meal +offer +choice full economy +passenger boarding +selection ...	749	7%
5	+option flying first +meal +recommend +time overall +offer comfortable plenty +entertainment +good +find +snack +movie ...	484	5%
6	economy business cathay premium pacific class first +lounge +class +seat +bed +service 'great experience' +experience +...	446	4%
7	+room +leg 'leg room' plenty +seat economy comfortable +entertainment +fly +meal 'inflight entertainment' +book +review +st...	385	4%
8	+great +experience great +'great service' 'great experience' +'great flight' +food +service +flight +fly +staff +crew comfortable 'o...	452	5%
9	+staff +help +airport +luggage +travel +connect +connecting flight' +day +miss +wait next helpful +check +time ground	540	5%
10	+staff ground +short +screen helpful +snack +friendly +movie +crew +drink professional +food +passenger +serve +cabin ...	1336	13%
11	lufthansa frankfurt class +short professional +nice first +connect +recommend +class economy +lounge business always pr...	175	2%
12	+entertainment +crew inflight +cabin +food +system 'inflight entertainment' +friendly +cabin crew' +'entertainment system' +mo...	878	9%
13	+class business +business class' +bed +seat +lounge economy first premium +service airways +selection +ticket attentive ...	761	8%
14	+flight +attendant +flight attendant' +nice +help professional +passenger +trip +friendly helpful attentive +food +service flying ...	541	5%

Final text cluster node: High SVD resolution with 105 dimensions

- 15 clusters generated

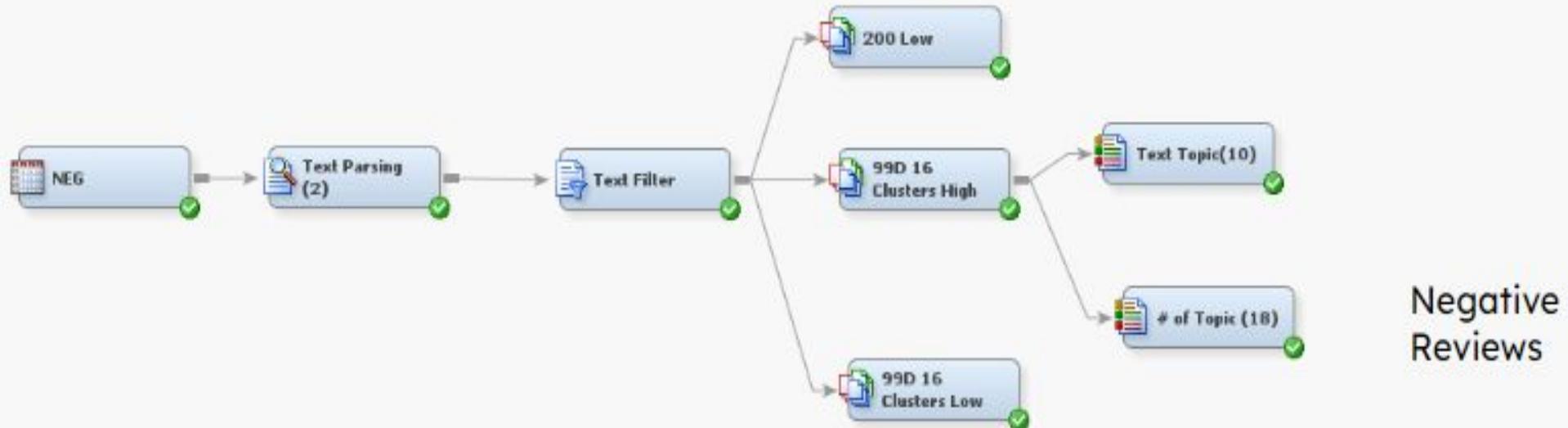
POSITIVE REVIEWS TEXT TOPICS

Topic	Number of Terms	# Docs
+entertainment,+cabin,+crew,+friendly,+cabin crew	346	1545
business,+class,+lounge,+business class,+bed	422	1627
+airport,+gate,+check,+luggage,+bag	574	1543
+serve,+breakfast,+minute,+drink,+hot	498	1568
economy,+leg,+room,leg room,+seat	501	1574

We further select 5 most crucial clusters and find below topics:

1. Friendly cabin crew
2. Business class lounge
3. Luggage and check bag in the airport
4. Serving breakfast and drinks
5. Leg room for seat

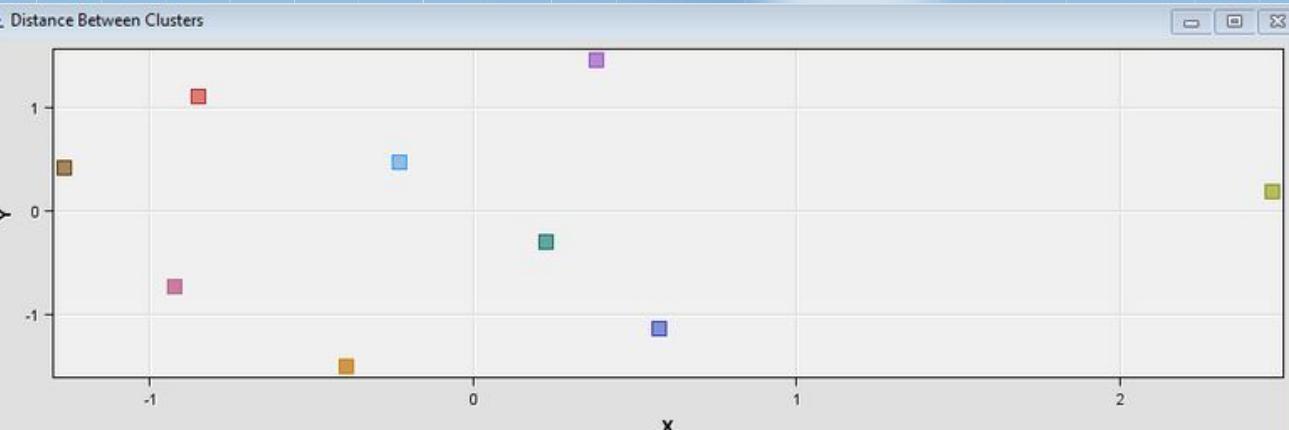
NEGATIVE REVIEWS



Negative
Reviews

NEGATIVE REVIEWS TEXT CLUSTERS

Distance Between Clusters



Cluster ID	Descriptive Terms	Frequency	Percentage
1	+seat +food +entertainment +meal +good +drink +offer +serve +cabin +poor +old +screen +system +small inflight	1698	16%
2	+flight +delay +time +airport +late +wait +miss +day +arrive +hotel +connect +leave +minute +cancel next	2157	20%
3	+crew +staff air +cabin +passenger china +'cabin crew' ground +food +serve +meal +rude +drink +experience +good	1222	11%
4	+service +customer +'customer service' +gate +cancel +line +rude +phone +help +day +wait +check +minute +experience worst	1153	11%
5	+bag +luggage +check +pay +day +arrive +charge +lose 'carry on' +baggage +claim +small +find +phone +cost	1044	10%
6	â en es - +con al el una si los +cost +charge +claim +lose +arrive	37	0%
7	+ticket +check boarding +pass +book +'boarding pass' +charge +pay +cost +baggage +cancel +airport +gate +phone +minute	1369	13%
8	+seat +attendant +leg +row +sit economy +room uncomfortable +pay +old +class first +small +fly +people	1244	12%
9	+class +business 'business class' airways economy etihad british +lounge +food +good +seat +cabin +serve +choice +old	863	8%

Final text cluster node: High SVD resolution with 99 dimensions

- 9 clusters generated

NEGATIVE REVIEWS TEXT TOPICS

Topic	Number of Terms	# Docs ▼
+hotel,+delay,+cancel,+delay,+connect	489	1784
+seat,+entertainment,+leg,+seat,+row	472	1746
+customer,+customer service,+service,+rude,worst	461	1685
+ticket,+book,+cancel,+customer,email	494	1635
+meal,+serve,+crew,+cabin,+drink	543	1577

We further select 5 most crucial clusters and find below topics:

1. Delay or cancellation of flight
2. Complaint of seats
3. Inappropriate customer service
4. Ticket booking issues
5. Meal serving by the cabin crews

CONCLUSION

- Cabin crew, seat and food serving are the topics that are frequently mentioned both in positive and negative comments
- Service are more important than the hardware
- Delay, cancellation or any other ticket-related issues of flight pose significant threat to the airline reputation



03

OBJECTIVE 2

To analyze how customers feel about airline service in a more detailed way

TO ANALYZE HOW CLIENTS FEEL ABOUT OUR SERVICE IN A MORE DETAILED WAY

Objective

To investigate text from positive and negative reviews to understand customers' perceptions toward different areas of the airline.

Methodology

Sentiment Analysis
Rule-Based Model
Hybrid Model

Data Input

Positive_Review
Negative_Review

DATA PREPARATION

- ❖ Data sampling
 - Use the random function in the python to randomly select 10,000 reviews of positive and negative for training analysis
 - Use the random function in the python to randomly select 1,000 reviews of positive and negative for test analysis
- ❖ Data output
 - Use python to output the each review become single text file

DATA PROCESSING

- SAS Sentiment Analysis Studio
- Building the Rule-Based Model to analysis the reviews data
- Set the different detailed way of airlines that customers may be concern
- Extract some subjective word that come from those airlines and classify them into definition, positive and negative.
- For Example: classify ‘seat’ as definition of cabin services , ‘professional’ , ‘comfortable’ as positive; ‘mess’ , ‘unfriendly’ as negative.

[Definitions](#) | [Positive](#) | [Negative](#) | [Neutral](#) |

	Type	Body	Weight
1	CLASSIFIER	english	1
2	CLASSIFIER	tables	1
3	CLASSIFIER	smile	1
4	CLASSIFIER	lavatory	1
5	CLASSIFIER	toilet	1
6	CLASSIFIER	bed	1
7	CLASSIFIER	blanket	1
8	CLASSIFIER	leg room	1
9	CLASSIFIER	smiles	1
10	CLASSIFIER	functioning service	1
11	CLASSIFIER	seat	1
12	CLASSIFIER	toilets	1
13	CLASSIFIER	bassinet	1
14	CLASSIFIER	seats	1
15	CLASSIFIER	table	1
16	CLASSIFIER	services	1
17	CLASSIFIER	stewards	1
18	CLASSIFIER	steward	1
19	CLASSIFIER		1

	Definitions	Positive	Negative	Neutral	
Type				Body	Weight
1	CLASSIFIER		no clear explanations		1
2	CLASSIFIER		slow		1
3	CLASSIFIER		worse		1
4	CLASSIFIER		mess		1
5	CLASSIFIER		lazy		1
6	CLASSIFIER		rude		1
7	CLASSIFIER		unhelpful		1
8	CLASSIFIER		unfriendly		1
9	CLASSIFIER		expensive		1
10	CLASSIFIER		disqualified		1
11	CLASSIFIER		unprofessional		1
12	CLASSIFIER		unacceptable		1
13	CLASSIFIER		no responsibility		1
14	CLASSIFIER		lies		1
15	CLASSIFIER		arrogant		1
16	CLASSIFIER		incompetent		1

	Definitions	Positive	Negative	Neutral	
Type				Body	Weight
1	CLASSIFIER		clear explanations		1
2	CLASSIFIER		excellent service		1
3	CLASSIFIER		comfortable seat		1
4	CLASSIFIER		clean		1
5	CLASSIFIER		comfortable		1
6	CLASSIFIER		polite		1
7	CLASSIFIER		clean		1
8	CLASSIFIER		professional		1
9	CLASSIFIER		attentive		1
10	CLASSIFIER		great		1
11	CLASSIFIER		pleasant		1
12	CLASSIFIER		efficient		1
13	CLASSIFIER		kind		1
14	CLASSIFIER		pretty		1
15	CLASSIFIER		decent		1
16	CLASSIFIER		spacious		1
17	CLASSIFIER		responsive		1
18	CLASSIFIER		comfort		1
19	CLASSIFIER		comfy		1
20	CLASSIFIER		resonable		1
21	CLASSIFIER		patient		1
22	CLASSIFIER		tidy		1
23	CLASSIFIER		fresh		1

RESULT

- ❖ For the Advanced Statistical model result
- ❖ The best model
 - Relative frequency and Chi Square
 - The overall precision is 89.76% , the positive precision is 85.91% and the negative precision is 93.33%

Statistical Model Configuration	
Training corpus	Airline_Review
Set percentage for training	80%
Solution	Bayes Method
Probability threshold	0.80
Text normalization model	Relative Frequency
Contextual extraction (optional)	
Runtime stop words (optional)	
Text Result	Graphical Result
With text normalization algorithm [Relative Frequency] and feature ranking algorithm [No Feature Ranking]:	
Overall precision: 88.41%	
Positive precision: 82.22%	
Negative precision: 94.16%	
With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Risk Ratio]:	
Overall precision: 81.56%	
Positive precision: 72.18%	
Negative precision: 90.27%	
With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Chi Square]:	
Overall precision: 89.76%	
Positive precision: 85.91%	
Negative precision: 93.33%	
With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Information Gain]:	
Overall precision: 89.71%	
Positive precision: 85.86%	
Negative precision: 93.28%	
BEST MODEL is Relative Frequency and Chi Square	

RESULT

Statistical model

Results for selected folder:
This directory is Negative
Overall accuracy is 89.90%;
Positive precision is 88.00%;
Negative precision is 91.80%.

Number of articles:1000
Number of positive articles:481
Number of negative articles:519
Number of neutral articles:0
Positive percent:48.10%.

Rule-Based model

Results for selected folder:
This directory is Negative
Overall accuracy is 89.90%;
Positive precision is 88.00%;
Negative precision is 91.80%.

Number of articles:1000
Number of positive articles:481
Number of negative articles:519
Number of neutral articles:0
Positive percent:48.10%.

Hybrid model

Results for selected folder:
This directory is Negative
Overall accuracy is 90.10%;
Positive precision is 88.00%;
Negative precision is 92.20%.

Number of articles:1000
Number of positive articles:479
Number of negative articles:521
Number of neutral articles:0
Positive percent:47.90%.

RESULT

1. Test the Positive reviews and Negative reviews by entering the text document
2. Classify the reviews whether positive or negative
3. Estimate that reviews belong to which detailed way

File: Test

First, I tried to upgrade. There were no seats left, but I still spent over 20 dollars to try and do so, because they're the only flagship airline I know that has a paid (and ridiculously expensive: 1 eur/min) customer service phone line. And it's obviously the only way you can upgrade, no option to do so from their website. The check-in was horrible, with a never ending line because it took them forever to check people in. In the end they had to ask help from the desks of another airline to cope with that. The flight experience was also horrible: the plane was old and filthy, the food was little and disgusting (I very rarely leave food on my plate, but some of that was inedible), the in-flight entertainment consisted of a screen just slightly bigger than my phone with a handful of movies that kept breaking for no apparent reason. The service on the plane was, guess what, horrible. Their stewards were borderline rude, throwing things on my tray when serving them, constantly hitting my head or shoulder because they didn't look where going. They were also arrogant and incompetent: when I asked for an immigration card, they told me I didn't need it anymore, and got annoyed as I insisted. In the end I didn't get one, and once I reached the immigration officer I had to go to the back of the queue because obviously I was right and they were wrong, and I did need an immigration card with my visa. I vowed in the past I wouldn't travel with Alitalia anymore. This time I couldn't avoid it, but they confirmed, once again if there was any need, that I made the right choice back then. I do hope they finally go out of business, as they absolutely deserve that.

Text Result | Graphical Result |

Overall Document: Test in hybrid-based model result is Negative
Probability to be positive is 0.02% with confidence 99.97%

CabinService: Test in rule-based model result is Negative
Probability to be positive is 16.49% with confidence 79.38%

GroundService: Test in rule-based model result is Negative
Probability to be positive is 11.64% with confidence 85.45%

FoodBeverage: Test in rule-based model result is Negative
Probability to be positive is 30.77% with confidence 61.54%

Entertainment: Test in rule-based model result is Negative
Probability to be positive is 40.00% with confidence 50.00%

24 matches:
6 matches for product definitions:
No 0 (42-46): CabinService : seats
No 1 (332-339): GroundService : checkin
No 2 (589-592): FoodBeverage : food

Product prominence information:
GroundService: Top 20%
CabinService: Top 20%
FoodBeverage: Bottom 80%
Entertainment: Bottom 80%

Product dominance information:
GroundService: In Passing
CabinService: In Passing
FoodBeverage: In Passing
Entertainment: In Passing

RECOMMENDATION

- ❖ After testing all positive reviews and negative reviews
- ❖ Result:

Food and Beverage:

- Offer a wider variety of menu options to cater to different dietary needs and preferences
- Improve the quality and taste of the food served on board
- Offer complimentary snacks and drinks in addition to meals to enhance the overall experience

Entertainment:

- Offer a wider selection of movies, TV shows, and music to cater to different interests and age groups
- Ensure that the in-flight entertainment system is functioning properly and is up-to-date
- Offer Wi-Fi connectivity and streaming services to allow passengers to access their own entertainment content

Ground Services:

- Improve the efficiency of the check-in process to reduce wait times and long queues
- Offer self-service kiosks for check-in and baggage drop-off to provide a quicker and more convenient experience
- Improve the cleanliness and upkeep of airport lounges and waiting areas

Cabin Services:

- Train staff to provide more friendly, attentive, and personalized service to passengers
- Ensure that cabins are clean, well-maintained, and comfortable
- Offer amenities such as blankets and eye masks to enhance passenger comfort during the flight.



04

OBJECTIVE 3

Identify the critical factor that affect customer's recommendation

OUR OBJECTIVE

Whether customers recommend the airline can directly reflect product quality and affect the company's revenue.

Therefore, we use different models to perform classification based on the customer's ratings for several categories of the flight.

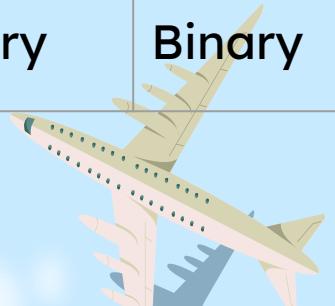
Finally, we want to find out the impact of those ratings on the customer's final recommendation.

HOW THIS OBJECTIVE IMPROVE THE INDUSTRY?

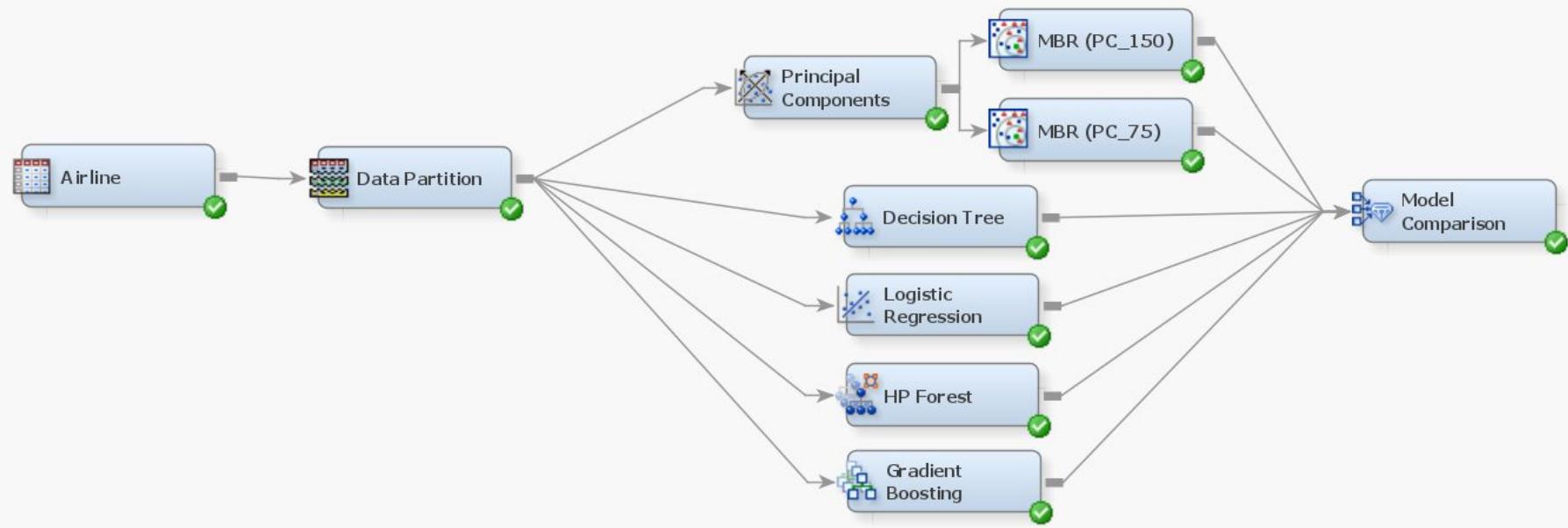
- 1.Improving customer satisfaction
- 2.Increasing profitability
- 3.Enhancing marketing efforts
- 4.Identifying areas for improvement

VARIABLE

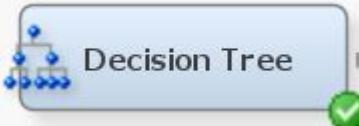
Input Variable	Type	Value
seat_comfort	Ordinal	(1 - 5)
cabin_service	Ordinal	(1 - 5)
food_bev	Ordinal	(1 - 5)
entertainment	Ordinal	(1 - 5)
ground_service	Ordinal	(1 - 5)
value_for_money	Ordinal	(1 - 5)
Target Variable	Type	Value
recommended_binary	Binary	(0 , 1)



SAS EM DIAGRAM



METHODOLOGY



Number of Leaves : 15

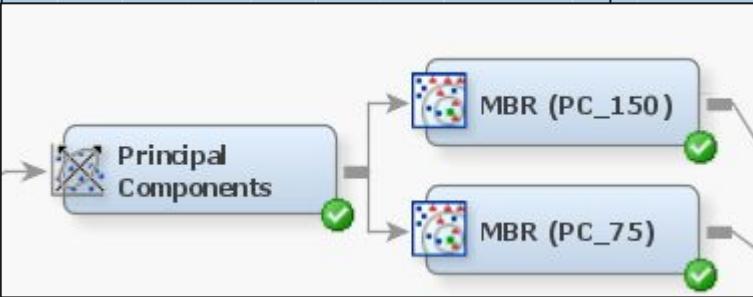
Assessment Measure :
Misclassification



Regression Type : Logistic
Selection Model : Stepwise
Selection Criterion : Validation
Misclassification



Maximum Number of Trees : 21
Variable Importance Method : Loss Reduction



Eigenvalue Cutoff (Cumulative): 0.99
Method: Scan

Model 1: Number of Neighbors=150
Model 2: Number of Neighbors=75



N Iterations : 49
Train Proportion : 100
Maximum Depth : 3
Split size : 100

MODEL COMPARISON

Selected Model	Model Description	Selection Criterion: Valid: Misclassification Rate	Train: Misclassification Rate
Y	HP Forest	0.056334941	0.055750402
	Gradient Boosting	0.056773345	0.056261873
	Logistic Regression	0.057430951	0.059111501
	Decision Tree	0.064007014	0.063203273
	MBR (PC_75)	0.069487067	0.066710507
	MBR (PC_150)	0.069487067	0.067441181

FINDING - VARIABLE IMPORTANCE

HP Forest (Valid: Gini Reduction)	
Variable name	Score
Value_for_money	0.15877
ground_service	0.07944
Cabin_service	0.07481
Seat_comfort	0.05386
Food_bev	0.02598
Entertainment	-0.00098

Based on the gini reduction results of the model, it appears that the **value_for_money is the most important factor** for predicting whether a customer will recommend an airline service, as it had the highest importance scores in the model.

The importance of "Ground_service" and "Cabin_service" are close.

"Seat_comfort" and "Food_bev" also had some importance, while "Entertainment" **did not appear to be an important factor in this model.**

RECOMMENDATION

Therefore, if companies decide to prioritize which area to focus on for improving customer recommendations, they should consider improving the perceived value for money of the service, as well as ground service and the quality of cabin service.

For example , Firms can offer more competitive pricing and incentives for customers, provide faster baggage handling and boarding processes or personalized cabin service.

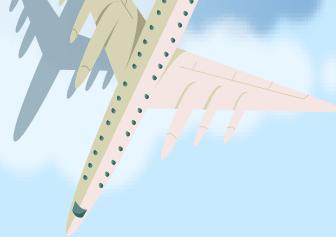
Improving seat comfort and food and beverage may also be useful, but may be less impact compare with other factors.



05

OBJECTIVE 4

Identify the aircraft components that consistently experience deflection



Objective 4

Aim:

- Identify the aircraft components that consistently experience defection.

By finding out those components, the aircraft maintenance department can put extra attention to those components so that we can minimize the occurrence of incidents

Methodology:

- 1. Clustering**
- 2. Topic extraction**

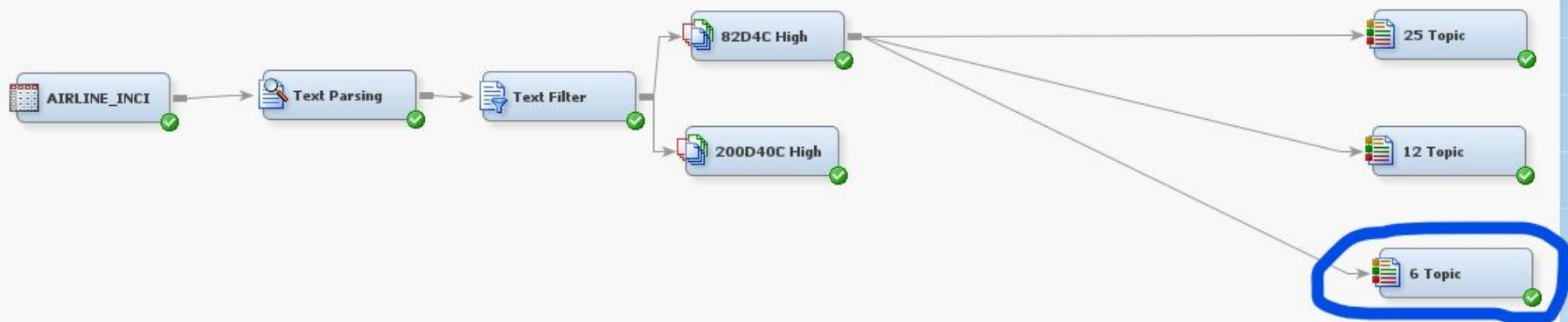
Data Input:

Report text from aircraft maintenance department

Report text contains details, like

- 1. Which component(s) occur error**
- 2. Procedures/ manual for repair**
- 3. File reference**

EM DIAGRAM



EXTRACTED COMPONENTS

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.158	0.012	+hinge,+elevator,+arm,+stabilizer,play	56	744
Multiple	2	0.141	0.012	emergency,battery,+supply,power,+unit	240	2449
Multiple	3	0.132	0.013	sta,+corrective action,corrective,+action,rev	376	2680
Multiple	4	0.121	0.014	+engine,+fault,yes,+ref,+task	457	2793
Multiple	5	0.122	0.013	+door,corrective,+corrective action,+action,+ship	363	3125
Multiple	6	0.131	0.013	floor,+panel,+base,base maintenance,maintenance	240	2307

1. Some of them are relatively not so important
EG) door, floor, panel
 - Affect customer's experience of the journey
 - Less likely to cause very serious accident
2. Engine
 - not so insightful since everyone knows the importance of the engine



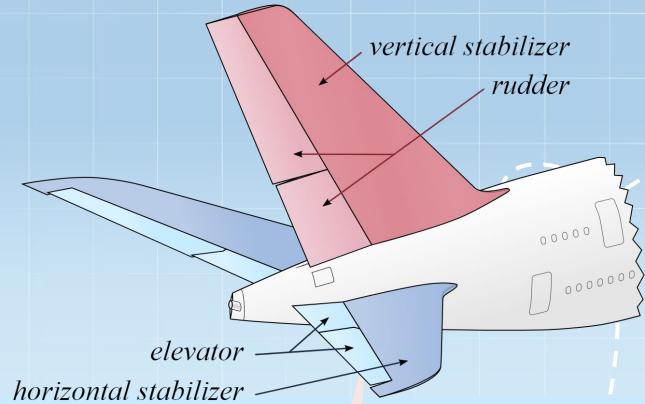
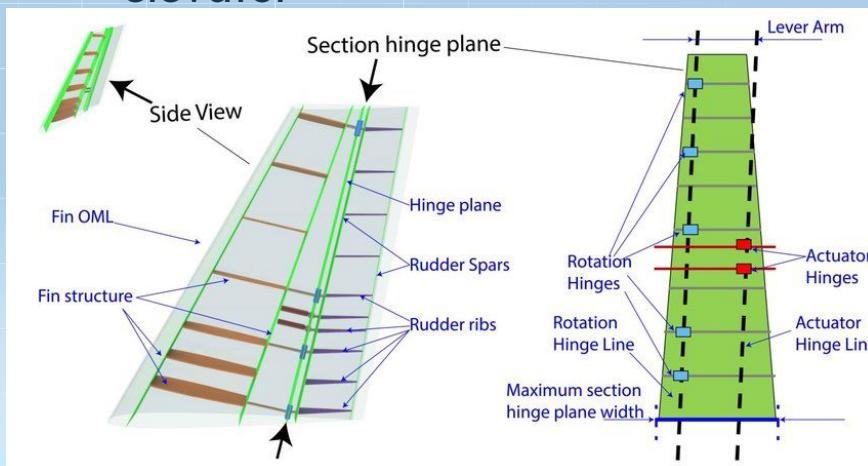
IMPORTANT FREQUENTLY DEFECTED COMPONENT

1. The back of the airplane

- Stabilizer i.e. horizontal and vertical stabilizer
- Hinges in the wing
- elevator

Topic

+hinge,+elevator,+arm,+stabilizer,play
emergency,battery,+supply,power,+unit .



2. Power system

- APU battery
- Used for power supply of the engine if main battery cannot be used

Ref: <http://787updates.newairplane.com/787-Electrical-Systems/Batteries-and-Advanced-Airplanes>



RECOMMENDATION

1. More attention during the regular maintenance
 - Detect any failures before they become major issues
2. Stock more inventories for replacement
 - Consistent error require more frequent replacement
3. Consider a back-up APU
 - More consideration should be put
 - APU is costly



THANKS!

APPENDIX - OBJECTIVE 1

APPENDIX - OBJECTIVE 1 - STOP LIST

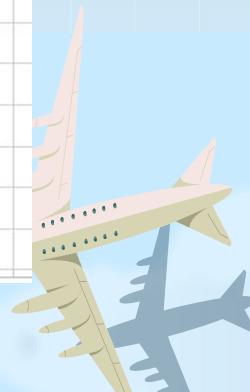
Positive & negative are sharing
same stop list

Total: 104 words

Types of words in the stop list:

1. Place name
2. Airline company name
3. Unit
4. Aircraft name
5. Aircraft number

A	B
Term	Role
plane	Noun
pm	Noun
aircraft	Noun
sunwing	Prop
min	Abbr
airplane	Noun
plane	Adj
tarmac	Noun
vueling	Prop
eventually	Adv
toronto	Prop
hrs	Noun
jetblue	Prop
didn	Noun
norwegian	Noun
frontier	Prop
people	Verb
jfk	Prop
ryanair	Prop
gatwick	Prop
barcelona	Prop
punta	Prop
dubai	Prop

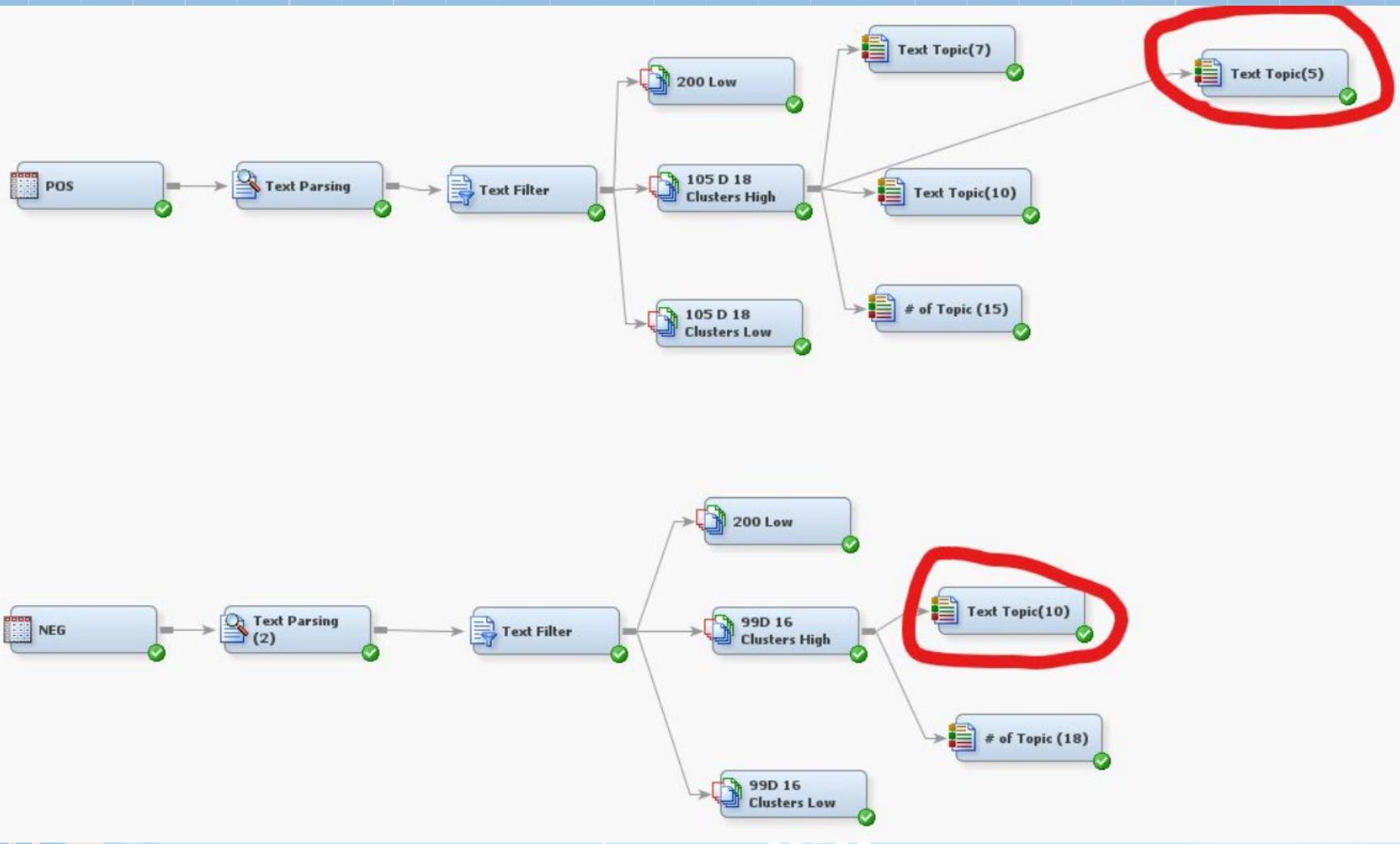


APPENDIX - OBJECTIVE 1 - SYNONYM LIST

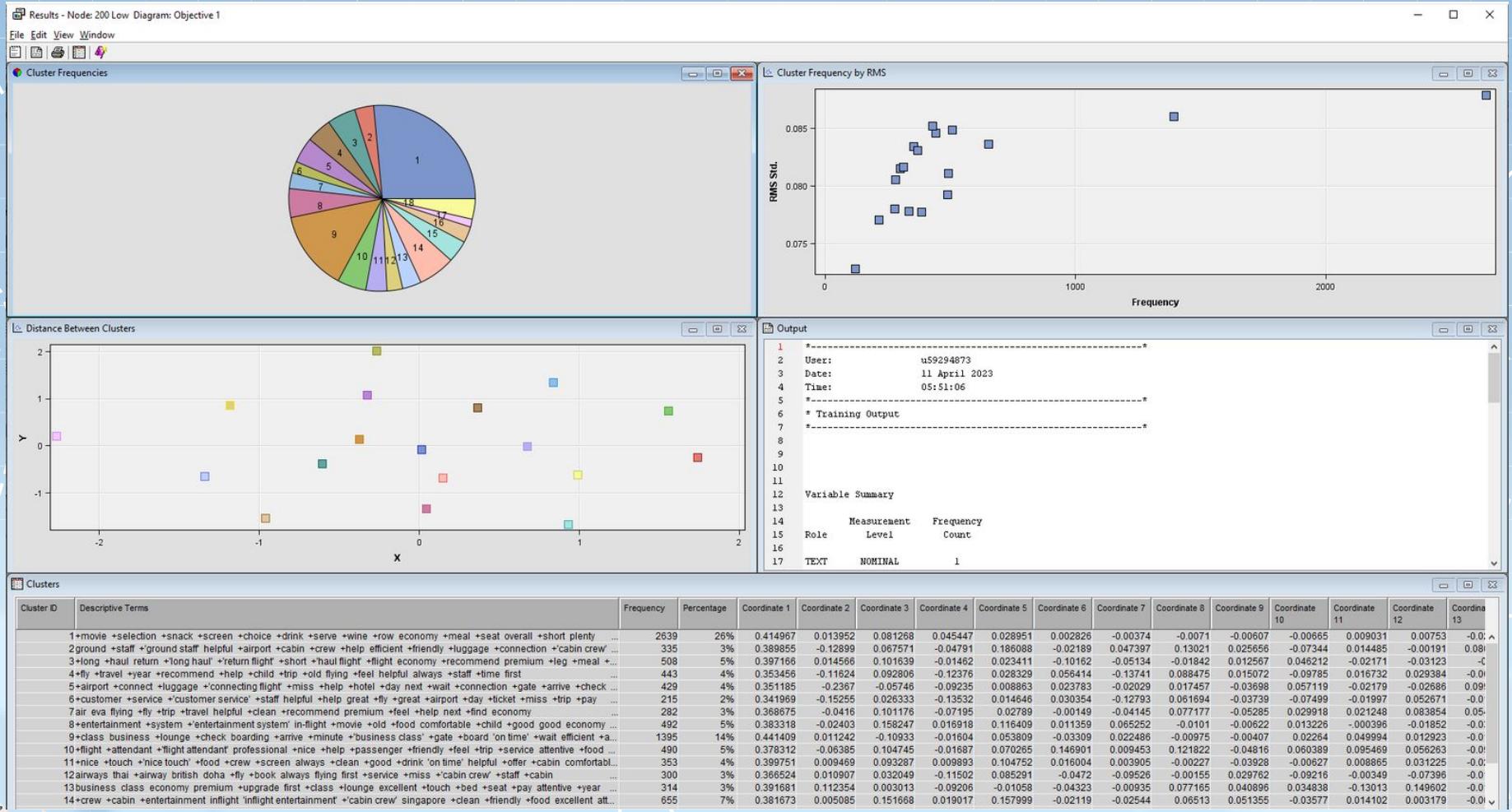
同義字-EMWS1.TextParsing_synonymDS

Child Term	Parent Term	Term Role	Parent Role
delay	overtime	Adj	Adj
nice	gentle	Adj	Adj
delay	postpone	Verb	Verb
old	elderly	Adj	Adl
nice	kind	Adj	Adj
issue	problem	Noun	Noun
rude	abusive	Adj	Adj
movie	film	Noun	Noun
rude	offensive	Adj	Adj
Nice	friendly	Adj	Adj
on time	not late	Adv	Adv
movie	cinema	Noun	Noun
recommend	advocate	Verb	Verb
meal	steak meal	Noun	Noun
comfortable	appropriate	Adj	Adj
old	aged	Adj	Adj

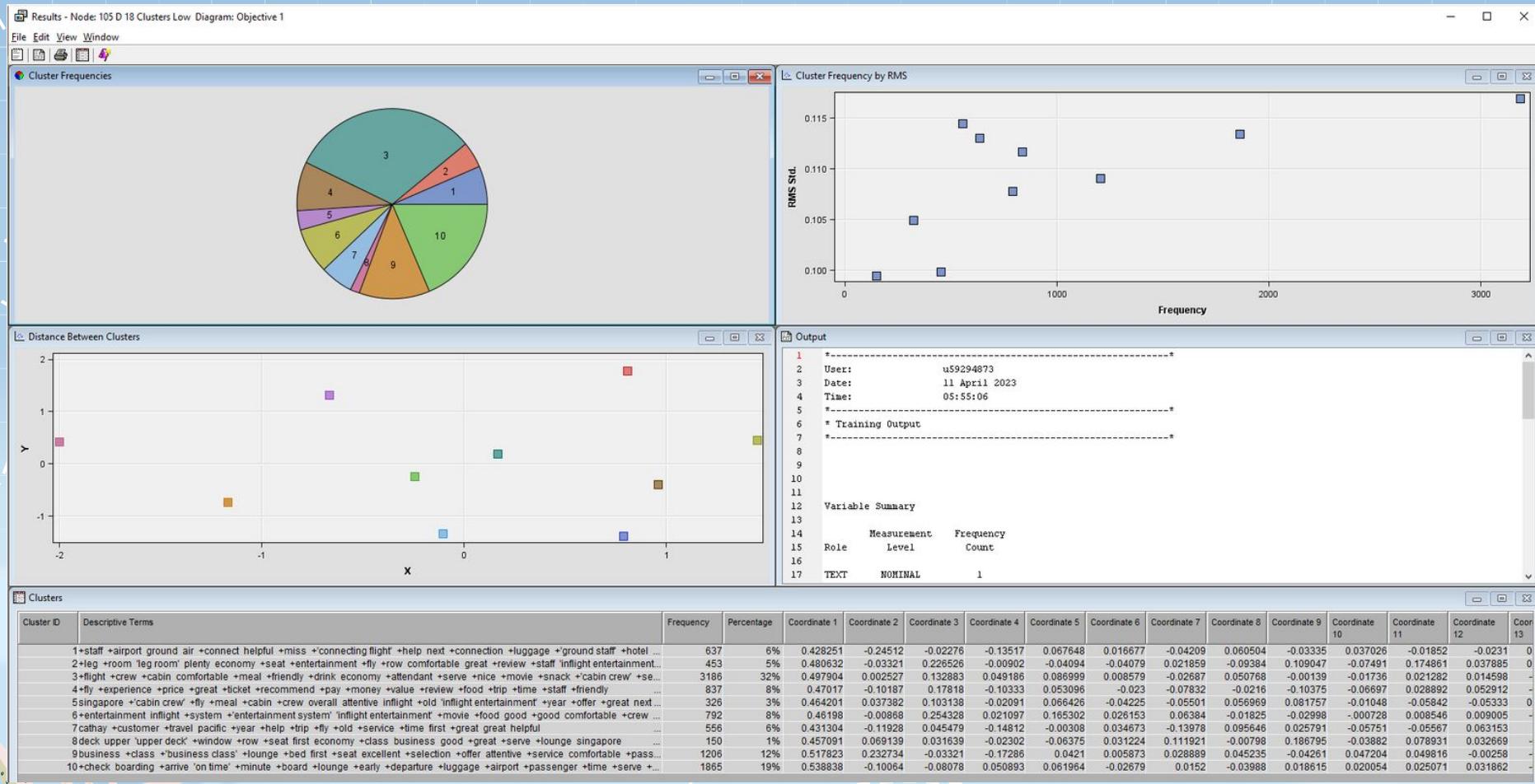
APPENDIX - OBJECTIVE 1 - SELECTED MODEL



APPENDIX - OBJECTIVE 1 - POSITIVE MODEL WITH 200 DIMENSIONS AND LOW RESOLUTIONS



APPENDIX - OBJECTIVE 1 - POSITIVE MODEL WITH 105 DIMENSIONS AND LOW RESOLUTIONS



APPENDIX - OBJECTIVE 1 - POSITIVE FINAL RESULT

Results - Node: Text Topic(5) Diagram: Objective 1

File Edit View Window

Number of Documents by Topics

Topic ID	# Docs
1	~1500
2	~1500
3	~1500
4	~1500
5	~1500

Topic Terms

Rank for Variable NUMDOCS

Rank	Value
1	17589

Number of Terms by Topics

Topic ID	Number of Terms
1	~350
2	~400
3	~550
4	~480
5	~500

Topics

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.135	0.013	+entertainment,+cabi...	346	1545
Multiple	2	0.100	0.013	business,+class,+iou...	422	1627
Multiple	3	0.096	0.014	+airport,+gate,+check...	574	1543
Multiple	4	0.090	0.014	+serve,+breakfast,+mi...	498	1568
Multiple	5	0.092	0.014	economy,+leg.+room,...	501	1574

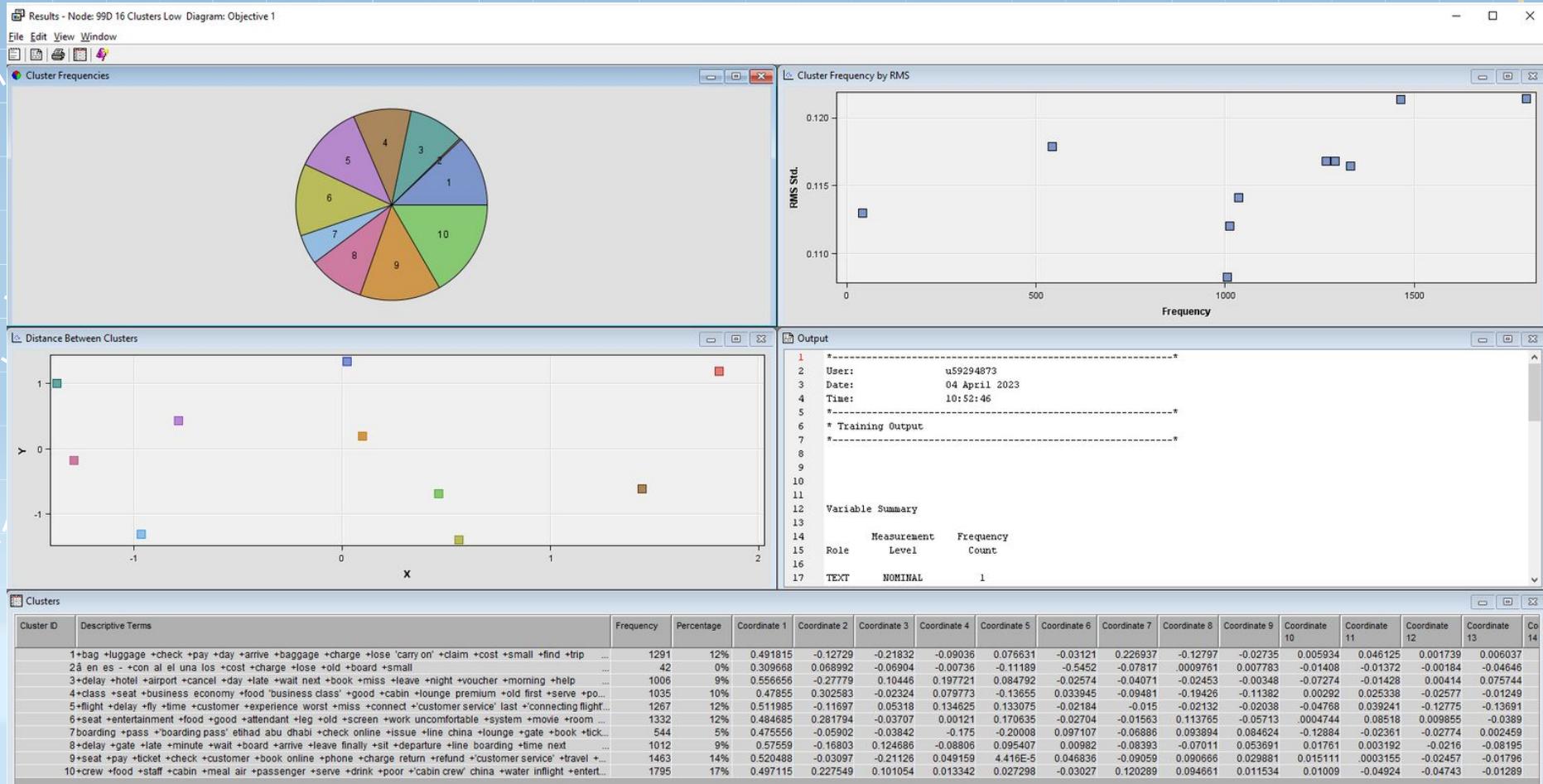
Output

```

1 -----
2 User: u59294873
3 Date: 11 April 2023
4 Time: 05:57:49
5 -----
6 * Training Output
7 -----
8
9
10
11
12 Variable Summary
13
14 Measurement Frequency
15 Role Level Count
16
17 TEXT NOMINAL 1

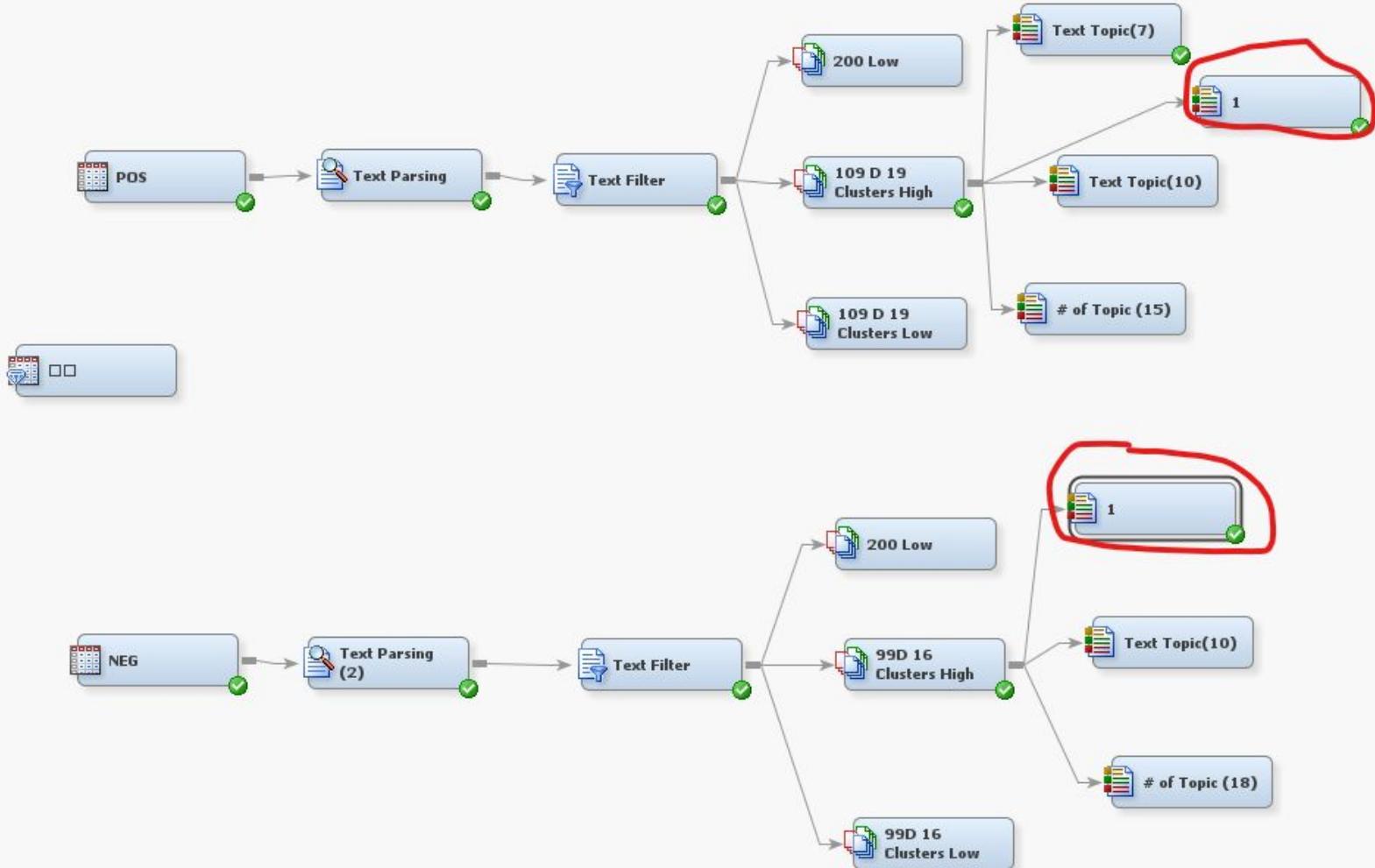
```

APPENDIX - OBJECTIVE 1 - NEGATIVE MODEL WITH 99 DIMENSIONS AND LOW RESOLUTIONS



APPENDIX - OBJECTIVE 2

APPENDIX - OBJECTIVE 2



Interactive Topic Viewer

File Edit



Topics

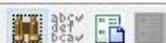
Topic	Category	Term Cutoff	Document Cutoff	Number of Terms	#
+flight,+seat,+good,+cabin,+crew	Multiple	0.014	0.172	629	1609

Terms

Topic Weight	+	Term	Role	# Docs	Freq
0.092	+	serve	Verb	1783	2262
0.091	+	passenger	Noun	1720	2356
0.09	+	drink	Noun	1821	2224
0.089	+	nice	Adj	1768	2229
0.088	+	leg	Noun	1533	2022
0.087	+	experience	Noun	1938	2285
0.087		on time	Adv	1819	2092
0.087	+	offer	Verb	1495	2012
0.085	+	lounge	Noun	1244	1916
0.083	+	great	Adj	1628	1970
0.079	+	attendant	Noun	1540	1860
0.079	+	choice	Noun	1432	1724

Interactive Topic Viewer

File Edit



Topics

Topic	Category	Term Cutoff	Document Cutoff	Number of T
+flight,+seat,+time,+airport,+service	Multiple	0.013	0.176	722

Terms

Topic Weight ▾	+	Term	Role	# Docs	Freq
0.063	+	want	Verb	1344	1597
0.063	+	sit	Verb	1251	1544
0.063	+	baggage	Noun	944	1499
0.061	+	long	Adj	1328	1581
0.061	+	poor	Adj	1295	1612
0.061	+	late	Adj	1143	1409
0.06	+	entertainment	Noun	1403	1668
0.06		worst	Adj	1367	1494
0.06	+	flight	Verb	1289	1458
0.059	+	seat	Verb	1100	1450
0.059	+	check	Noun	1030	1344
0.058	+	problem	Noun	1123	1382

```
In [46]: #500 for positive and neg
df_ppp = df_positive['review_text'].reset_index()
lst_pos=[]

for i in range(500):
    lst_pos.append(df_ppp['review_text'][random.randint(0, len(df_positive))])

lst_neg=[]
df_nnn = df_negative['review_text'].reset_index()

for i in range(500):
    lst_neg.append(df_nnn['review_text'][random.randint(0, len(df_negative))])
```

```
In [47]: for x, i in enumerate (lst_pos):
    outfile = open('test/pos/'+'pos_'+str(x+1)+'.txt','w', encoding="utf-8")
    outfile.write(str(i))
    outfile.close()

for x, i in enumerate (lst_neg):
    outfile = open('test/neg/'+'neg_'+str(x+1)+'.txt','w', encoding="utf-8")
    outfile.write(str(i))
    outfile.close()
```

```
#10,000 for train
df_ppp = df_positive['review_text'].reset_index()
lst_pos=[]

for i in range(5000):
    lst_pos.append(df_ppp['review_text'][random.randint(0, len(df_positive))])

lst_neg=[]
df_nnn = df_negative['review_text'].reset_index()

for i in range(5000):
    lst_neg.append(df_nnn['review_text'][random.randint(0, len(df_negative))])

for x, i in enumerate (lst_pos):
    outfile = open('test/pos/'+'pos_'+str(x+1)+'.txt','w', encoding="utf-8")
    outfile.write(str(i))
    outfile.close()

for x, i in enumerate (lst_neg):
    outfile = open('test/neg/'+'neg_'+str(x+1)+'.txt','w', encoding="utf-8")
    outfile.write(str(i))
    outfile.close()
```



Corpora
Statistical
Rule
Test

- Tonal Keyword
- Intermediate Entities
- Products
 - FoodBeverage
 - Entertainment
 - CabinService
 - GroundService

Rules

Search Rules

Positive | Negative | Neutral |

All types

Type

Body

1	CLASSIFIER	polite
2	CLASSIFIER	clean
3	CLASSIFIER	professional
4	CLASSIFIER	attentive
5	CLASSIFIER	great
6	CLASSIFIER	pleasant
7	CLASSIFIER	efficient
8	CLASSIFIER	delicious
9	CLASSIFIER	kind
10	CLASSIFIER	well
11	CLASSIFIER	smooth
12	CLASSIFIER	fantastic
13	CLASSIFIER	satisfy
14	CLASSIFIER	pretty
15	CLASSIFIER	decent
16	CLASSIFIER	spacious
17	CLASSIFIER	clean
18	CLASSIFIER	modern
19	CLASSIFIER	amaze
20	CLASSIFIER	fast
21	CLASSIFIER	enjoyable
22	CLASSIFIER	update
23	CLASSIFIER	responsive
24	CLASSIFIER	comfort
25	CLASSIFIER	comfy
26	CLASSIFIER	quick
27	CLASSIFIER	resonable
28	CLASSIFIER	impressive
29	CLASSIFIER	patient
30	CLASSIFIER	amazing
31	CLASSIFIER	generous
32	CLASSIFIER	tidy
33	CLASSIFIER	good quality
34	CLASSIFIER	fresh
35	CLASSIFIER	free
36	CLASSIFIER	excellent
37	CLASSIFIER	Interesting
38	CLASSIFIER	newer
39	CLASSIFIER	wonderful
40	CLASSIFIER	good
41	CLASSIFIER	safety



	Rules
Corpora	Tonal Keyword
	Intermediate Entities
Products	FoodBeverage
	Entertainment
	CabinService
	GroundService

	Type	Body	Weight
1	CLASSIFIER	crowded	1
2	CLASSIFIER	delayed	1
3	CLASSIFIER	hard	1
4	CLASSIFIER	slow	1
5	CLASSIFIER	worse	1
6	CLASSIFIER	mess	1
7	CLASSIFIER	useless	1
8	CLASSIFIER	disaster	1
9	CLASSIFIER	cancel	1
10	CLASSIFIER	lazy	1
11	CLASSIFIER	poor	1
12	CLASSIFIER	delay	1
13	CLASSIFIER	loss	1
14	CLASSIFIER	Delays	1
15	CLASSIFIER	rude	1
16	CLASSIFIER	unhelpful	1
17	CLASSIFIER	disastrous	1
18	CLASSIFIER	unfriendly	1
19	CLASSIFIER	lost	1
20	CLASSIFIER	Terrible	1
21	CLASSIFIER	expensive	1
22	CLASSIFIER	How come	1
23	CLASSIFIER	fault	1
24	CLASSIFIER	lies	1
25	CLASSIFIER	disqualified	1
26	CLASSIFIER	tightest	1
27	CLASSIFIER	unprofessional	1
28	CLASSIFIER	yelling	1

SAS Sentiment Analysis Studio - Airline

File Edit View Build Help



Corpora

- Tonal Keyword
- Intermediate Entities
- Products
 - FoodBeverage
 - Entertainment
 - CabinService
 - GroundService

Statistical

Rule

Test

Rules

Search Rules

All types

Positive

Negative

Neutral

Type

Body

Weight

	Type	Body	Weight
1	CLASSIFIER	average	1
2	CLASSIFIER	vast	1
3	CLASSIFIER	easy	1
4	CLASSIFIER	cold	1
5	CLASSIFIER	hot	1
6	CLASSIFIER	big	1
7	CLASSIFIER	cheap	1
8	CLASSIFIER	overbooked	1
9	CLASSIFIER	Never	1
10	CLASSIFIER	easier	1
11	CLASSIFIER	fault	1

amenities

F&B



Search Rules		All types
Type	Body	Weight
1 CLASSIFIER	wine	1
2 CLASSIFIER	meal	1
3 CLASSIFIER	coffee	1
4 CLASSIFIER	snack	1
5 CLASSIFIER	sandwich	1
6 CLASSIFIER	beer	1
7 CLASSIFIER	Breakfast	1
8 CLASSIFIER	alcohol	1
9 CLASSIFIER	AVML	1
10 CLASSIFIER	cheese	1
11 CLASSIFIER	vegetables	1
12 CLASSIFIER	steak	1
13 CLASSIFIER	steak meal	1
14 CLASSIFIER	dinner	1
15 CLASSIFIER	drink	1
16 CLASSIFIER	lunch	1
17 CLASSIFIER	tea	1
18 CLASSIFIER	water	1
19 CLASSIFIER	food	1
20 CLASSIFIER	drinks	1

Search Rules | All types

Definitions Positive | Negative | Neutral

Type	Body	Weight
1 CLASSIFIER	hot	1
2 CLASSIFIER	fantastic	1
3 CLASSIFIER	delicious	1
4 CLASSIFIER	decent	1
5 CLASSIFIER	Lovely	1
6 CLASSIFIER	wonderful	1
7 CLASSIFIER	good quality	1
8 CLASSIFIER		1



Search Rules

Definitions | Positive | **Negative** | Neutral |

Type

1	CLASSIFIER	cold food
2	CLASSIFIER	tasteless
3	CLASSIFIER	uneatable
4	CLASSIFIER	little
5	CLASSIFIER	disgusting
6	CLASSIFIER	

ENTERTAINMENT

	Definitions	Positive	Negative	Neutral	
	Type	Body			Weight
1	CLASSIFIER	movie			1
2	CLASSIFIER	tv			1
3	CLASSIFIER	wifi			1
4	CLASSIFIER	headphones			1
5	CLASSIFIER	book			1
6	CLASSIFIER	film			1
7	CLASSIFIER	magazine			1
8	CLASSIFIER	game			1
9	CLASSIFIER	movie selection			1
10	CLASSIFIER	video			1
11	CLASSIFIER	wi-fi			1
12	CLASSIFIER	touch screen			1
13	CLASSIFIER	tv show			1
14	CLASSIFIER	new movie			1
15	CLASSIFIER	television screen			1
16	CLASSIFIER	entertainment system			1
17	CLASSIFIER	entertainment			1
18	CLASSIFIER				1

Search Rules

All types

Definitions | Positive | **Negative** | Neutral

Type	Body	Weight
1 CLASSIFIER	slow	1
2 CLASSIFIER	handful of movies	1
3 CLASSIFIER	useless	1
4 CLASSIFIER	poor	1

Search Rules

All types

	Type	Body	Weight
1	CLASSIFIER	smooth	1
2	CLASSIFIER	satisfy	1
3	CLASSIFIER	pretty	1
4	CLASSIFIER	decent	1
5	CLASSIFIER	modern	1
6	CLASSIFIER	fast	1
7	CLASSIFIER	update	1
8	CLASSIFIER	quick	1
9	CLASSIFIER	excellent	1
10	CLASSIFIER	vigorous	1
11	CLASSIFIER	lastest movie	1

CABIN SERVICE

Search Rules | All types

Definitions | Positive | Negative | Neutral |

Type	Body	Weight
1	CLASSIFIER english	1
2	CLASSIFIER tables	1
3	CLASSIFIER smile	1
4	CLASSIFIER lavatory	1
5	CLASSIFIER toilet	1
6	CLASSIFIER bed	1
7	CLASSIFIER blanket	1
8	CLASSIFIER leg room	1
9	CLASSIFIER smiles	1
10	CLASSIFIER functioning service	1
11	CLASSIFIER seat	1
12	CLASSIFIER toilets	1
13	CLASSIFIER bassinet	1
14	CLASSIFIER seats	1
15	CLASSIFIER table	1
16	CLASSIFIER services	1
17	CLASSIFIER stewards	1
18	CLASSIFIER steward	1
19	CLASSIFIER	1

Definitions | Positive | **Negative** | Neutral |

Type	Body	Weight
1 CLASSIFIER	no clear explanations	1
2 CLASSIFIER	slow	1
3 CLASSIFIER	worse	1
4 CLASSIFIER	mess	1
5 CLASSIFIER	lazy	1
6 CLASSIFIER	rude	1
7 CLASSIFIER	unhelpful	1
8 CLASSIFIER	unfriendly	1
9 CLASSIFIER	expensive	1
10 CLASSIFIER	disqualified	1
11 CLASSIFIER	unprofessional	1
12 CLASSIFIER	unacceptable	1
13 CLASSIFIER	no responsibility	1
14 CLASSIFIER	lies	1
15 CLASSIFIER	arrogant	1
16 CLASSIFIER	incompetent	1
17 CLASSIFIER	throwing things	1

Search Rules

All types

Type	Body	Weight
1 CLASSIFIER	clear explanations	1
2 CLASSIFIER	excellent service	1
3 CLASSIFIER	comfortable seat	1
4 CLASSIFIER	clean	1
5 CLASSIFIER	comfortable	1
6 CLASSIFIER	polite	1
7 CLASSIFIER	clean	1
8 CLASSIFIER	professional	1
9 CLASSIFIER	attentive	1
10 CLASSIFIER	great	1
11 CLASSIFIER	pleasant	1
12 CLASSIFIER	efficient	1
13 CLASSIFIER	kind	1
14 CLASSIFIER	pretty	1
15 CLASSIFIER	decent	1
16 CLASSIFIER	spacious	1
17 CLASSIFIER	responsive	1
18 CLASSIFIER	comfort	1
19 CLASSIFIER	comfy	1
20 CLASSIFIER	resonable	1
21 CLASSIFIER	patient	1
22 CLASSIFIER	tidy	1
23 CLASSIFIER	fresh	1
24 CLASSIFIER	friendly	1
25 CLASSIFIER	comfortable	1

GROUND SERVICE

Definitions		Positive	Negative	Neutral	
	Type			Body	Weight
1	CLASSIFIER	check-in			1
2	CLASSIFIER	english			1
3	CLASSIFIER	lounge			1

Search Rules

All types

	Type	Body	Weight
1	CLASSIFIER	no clear explanations	1
2	CLASSIFIER	lost baggage	1
3	CLASSIFIER	slow	1
4	CLASSIFIER	worse	1
5	CLASSIFIER	mess	1
6	CLASSIFIER	lazy	1
7	CLASSIFIER	rude	1
8	CLASSIFIER	unhelpful	1
9	CLASSIFIER	unfriendly	1
10	CLASSIFIER	expensive	1
11	CLASSIFIER	disqualified	1
12	CLASSIFIER	unprofessional	1
13	CLASSIFIER	unacceptable	1
14	CLASSIFIER	no responsibility	1
15	CLASSIFIER	lies	1
16	CLASSIFIER	arrogant	1
17	CLASSIFIER	incompetent	1
18	CLASSIFIER	never ending line	1
19	CLASSIFIER		1
20	CLASSIFIER		1
21	CLASSIFIER		1

Definitions		Positive	Negative	Neutral	
	Type	Body			Weight
1	CLASSIFIER	clear explanations			1
2	CLASSIFIER	smiles			1
3	CLASSIFIER	smile			1
4	CLASSIFIER	excellent service			1
5	CLASSIFIER				1
6	CLASSIFIER				1

SIMPLE TEST

SAS Sentiment Analysis Studio - Airline

File Edit View Build Help

Corpora Statistical Models

Advanced Simple

Statistical Model Configuration

Training corpus: Airline_Review
Set percentage for training: 80%

Best mode

Text Result | Graphical Result |

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [No Feature Ranking]:
Overall precision: 88.41%
Positive precision: 82.22%
Negative precision: 94.16%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Risk Ratio]:
Overall precision: 81.56%
Positive precision: 72.18%
Negative precision: 90.27%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Chi Square]:
Overall precision: 89.76%
Positive precision: 85.91%
Negative precision: 93.33%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Information Gain]:
Overall precision: 89.71%
Positive precision: 85.86%
Negative precision: 93.28%

BEST MODEL is Relative Frequency and Chi Square

ADVANCED 0.2

Statistical Models Statistical Model Configuration

Advanced Simple

Training corpus: Airline_Review
Set percentage for training: 80%
Solution: Bayes Method
Probability threshold: 0.20
Text normalization model: Relative Frequency
Contextual extraction (optional):
Runtime stop words (optional):

[Text Result](#) | [Graphical Result](#)

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [No Feature Ranking]:
Overall precision: 88.41%
Positive precision: 82.22%
Negative precision: 94.16%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Risk Ratio]:
Overall precision: 81.56%
Positive precision: 72.18%
Negative precision: 90.27%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Chi Square]:
Overall precision: 89.76%
Positive precision: 85.91%
Negative precision: 93.33%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Information Gain]:
Overall precision: 99.71%
Positive precision: 85.86%
Negative precision: 93.28%

BEST MODEL is Relative Frequency and Chi Square

ADVANCED 0.8

SAS Sentiment Analysis Studio - Airline

File Edit View Build Help

Corpora Statistical Rule Test

Advanced Simple

Statistical Models

Statistical Model Configuration

Training corpus: Airline_Review
Set percentage for training: 80%

Solution: Bayes Method
Probability threshold: 0.80
Text normalization model: Relative Frequency
Contextual extraction (optional):
Runtime stop words (optional):

Text Result | Graphical Result

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [No Feature Ranking]:
Overall precision: 88.41%
Positive precision: 82.22%
Negative precision: 94.16%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Risk Ratio]:
Overall precision: 81.56%
Positive precision: 72.18%
Negative precision: 90.27%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Chi Square]:
Overall precision: 99.76%
Positive precision: 85.91%
Negative precision: 93.33%

With text normalization algorithm [Relative Frequency] and feature ranking algorithm [Information Gain]:
Overall precision: 89.71%
Positive precision: 85.86%
Negative precision: 93.28%

BEST MODEL is Relative Frequency and Chi Square

HYBRID 0.2

File Edit View Build Help

Test Data File: C:/Users/mhyan2/Desktop/Sent/Sentiment/negative/neg_2.txt Test

Corpora Statistical Rule Test

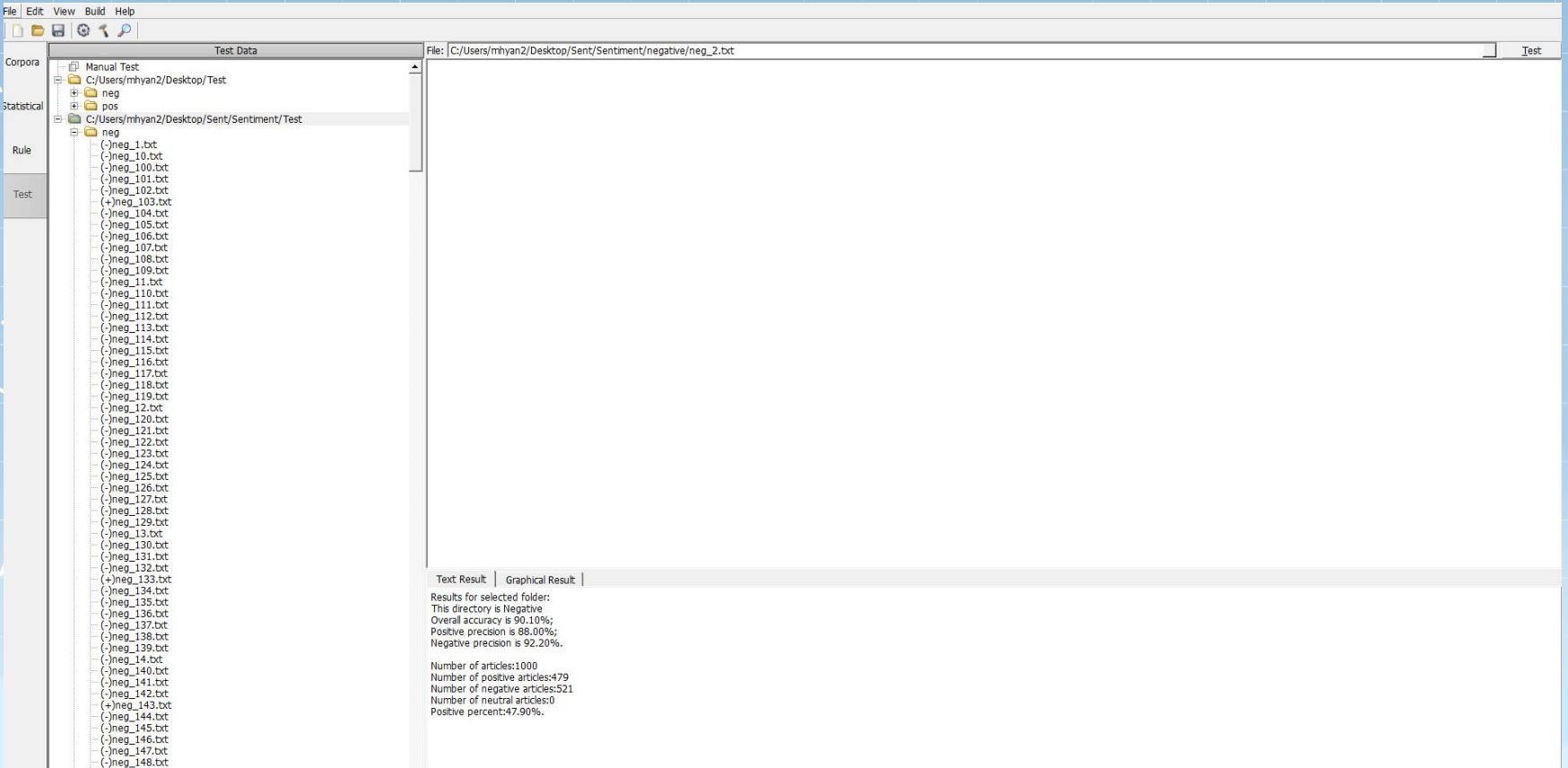
Manual Test
C:/Users/mhyan2/Desktop/Test
neg
pos

C:/Users/mhyan2/Desktop/Sent/Sentiment/Test
neg
(-)neg_1.txt
(-)neg_10.txt
(+)-neg_100.txt
(+)-neg_101.txt
(-)neg_102.txt
(+)-neg_103.txt
(-)neg_104.txt
(-)neg_105.txt
(-)neg_106.txt
(-)neg_107.txt
(-)neg_108.txt
(-)neg_109.txt
(-)neg_110.txt
(-)neg_111.txt
(-)neg_112.txt
(-)neg_113.txt
(-)neg_114.txt
(+)-neg_115.txt
(+)-neg_116.txt
(-)neg_117.txt
(-)neg_118.txt
(-)neg_119.txt
(-)neg_12.txt
(-)neg_120.txt
(-)neg_121.txt
(-)neg_122.txt
(-)neg_123.txt
(-)neg_124.txt
(-)neg_125.txt
(-)neg_126.txt
(-)neg_127.txt
(-)neg_128.txt
(-)neg_129.txt
(-)neg_13.txt
(-)neg_130.txt
(-)neg_131.txt
(-)neg_132.txt
(+)-neg_133.txt
(-)neg_134.txt
(-)neg_135.txt
(-)neg_136.txt
(+)-neg_137.txt
(-)neg_138.txt
(-)neg_139.txt
(-)neg_14.txt
(-)neg_140.txt
(-)neg_141.txt
(-)neg_142.txt
(+)-neg_143.txt
(-)neg_144.txt
(-)neg_145.txt
(-)neg_146.txt
(-)neg_147.txt
(-)neg_148.txt
(-)neg_149.txt
(-)neg_150.txt
(+)-neg_151.txt

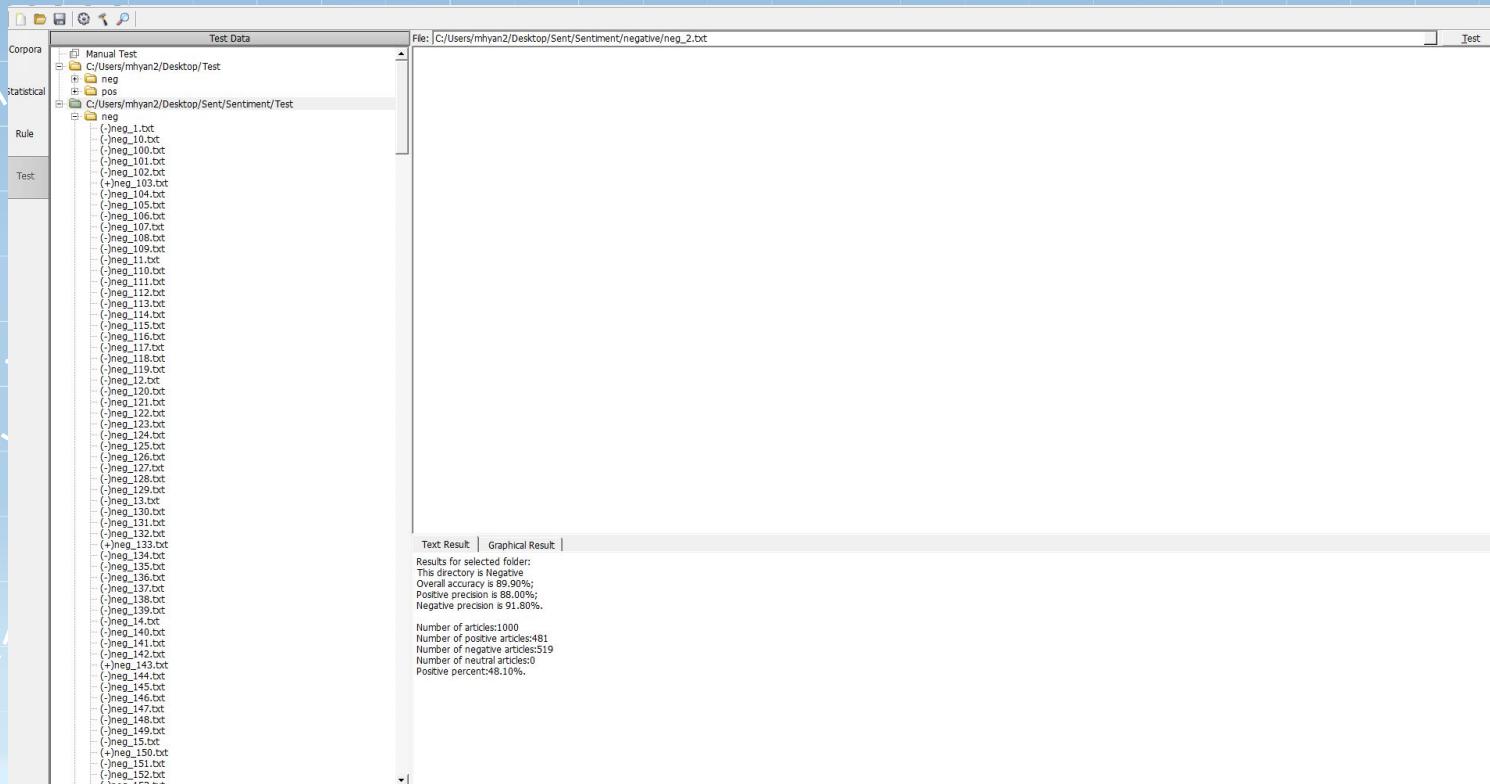
Text Result | Graphical Result |
Results for selected folder:
This directory is Positive
Overall accuracy is 87.90%;
Positive precision is 93.00%;
Negative precision is 82.80%.

Number of articles:1000
Number of positive articles:551
Number of negative articles:449
Number of neutral articles:0
Positive percent:55.10%.

HYBRID 0.8



STAT TEST 0.8



STAT TEST 0.2

File Edit View Build Help

Corpora Statistical Rule Test

Test Data File: C:/Users/mhyan2/Desktop/Sent/Sentiment/negative/neg_2.txt Test

Manual Test
C:/Users/mhyan2/Desktop/Test
neg
pos
C:/Users/mhyan2/Desktop/Sent/Sentiment/Test
neg
(-)neg_1.txt
(-)neg_2.txt
(-)neg_100.txt
(+)-neg_101.txt
(-)neg_102.txt
(+)-neg_103.txt
(-)neg_104.txt
(-)neg_105.txt
(-)neg_106.txt
(-)neg_107.txt
(-)neg_108.txt
(-)neg_109.txt
(-)neg_111.txt
(-)neg_110.txt
(-)neg_111.txt
(-)neg_112.txt
(-)neg_113.txt
(-)neg_114.txt
(+)-neg_115.txt
(+)-neg_116.txt
(-)neg_117.txt
(-)neg_118.txt
(-)neg_119.txt
(-)neg_120.txt
(-)neg_121.txt
(-)neg_122.txt
(-)neg_123.txt
(-)neg_124.txt
(-)neg_125.txt
(-)neg_126.txt
(-)neg_127.txt
(-)neg_128.txt
(-)neg_129.txt
(-)neg_130.txt
(-)neg_131.txt
(-)neg_132.txt
(+)-neg_133.txt
(-)neg_134.txt
(-)neg_135.txt
(-)neg_136.txt
(+)-neg_137.txt
(-)neg_138.txt
(-)neg_139.txt
(-)neg_140.txt
(-)neg_141.txt
(-)neg_142.txt
(+)-neg_143.txt
(-)neg_144.txt
(-)neg_145.txt
(+)-neg_146.txt
(-)neg_147.txt
(-)neg_148.txt
(-)neg_149.txt
(-)neg_150.txt
(+)-neg_151.txt

Text Result | Graphical Result |

Results for selected folder:
This directory is Positive.
Overall accuracy is 87.90%;
Positive precision is 93.00%;
Negative precision is 82.80%.

Number of articles:1000
Number of positive articles:551
Number of negative articles:449
Number of neutral articles:0
Positive percent:55.10%.

RULE BASE 0.2

File Edit View Build Help

Corpora Statistical Rule Test

Test Data File: C:/Users/mhyan2/Desktop/Sent/Sentiment/negative/neg_2.txt Test

Manual Test
C:/Users/mhyan2/Desktop/Test
neg
pos
C:/Users/mhyan2/Desktop/Sent/Sentiment/Test
neg
(-)neg_1.txt
(-)neg_2.txt
(-)neg_100.txt
(+)-neg_101.txt
(-)neg_102.txt
(+)-neg_103.txt
(-)neg_104.txt
(-)neg_105.txt
(-)neg_106.txt
(-)neg_107.txt
(-)neg_108.txt
(-)neg_109.txt
(-)neg_111.txt
(-)neg_110.txt
(-)neg_111.txt
(-)neg_112.txt
(-)neg_113.txt
(-)neg_114.txt
(+)-neg_115.txt
(+)-neg_116.txt
(-)neg_117.txt
(-)neg_118.txt
(-)neg_119.txt
(-)neg_120.txt
(-)neg_121.txt
(-)neg_122.txt
(-)neg_123.txt
(-)neg_124.txt
(-)neg_125.txt
(-)neg_126.txt
(-)neg_127.txt
(-)neg_128.txt
(-)neg_129.txt
(-)neg_130.txt
(-)neg_131.txt
(-)neg_132.txt
(+)-neg_133.txt
(-)neg_134.txt
(-)neg_135.txt
(-)neg_136.txt
(+)-neg_137.txt
(-)neg_138.txt
(-)neg_139.txt
(-)neg_140.txt
(-)neg_141.txt
(-)neg_142.txt
(+)-neg_143.txt
(-)neg_144.txt
(-)neg_145.txt
(+)-neg_146.txt
(-)neg_147.txt
(-)neg_148.txt
(-)neg_149.txt
(-)neg_150.txt
(+)-neg_151.txt

Text Result | Graphical Result |

Results for selected folder:
This directory is Positive.
Overall accuracy is 87.90%;
Positive precision is 93.00%;
Negative precision is 82.80%.

Number of articles:1000
Number of positive articles:551
Number of negative articles:449
Number of neutral articles:0
Positive percent:55.10%

RULE BASE 0.8

File Edit View Build Help

Corpora Statistical Rule Test

Test Data File: C:/Users/mhyan2/Desktop/Sent/Sentiment/negative/neg_2.txt Test

Manual Test
C:/Users/mhyan2/Desktop/Test
neg
pos
C:/Users/mhyan2/Desktop/Sent/Sentiment/Test
neg
(-)neg_1.txt
(-)neg_2.txt
(-)neg_100.txt
(+)-neg_101.txt
(-)neg_102.txt
(+)-neg_103.txt
(-)neg_104.txt
(-)neg_105.txt
(-)neg_106.txt
(-)neg_107.txt
(-)neg_108.txt
(-)neg_109.txt
(-)neg_111.txt
(-)neg_110.txt
(-)neg_111.txt
(-)neg_112.txt
(-)neg_113.txt
(-)neg_114.txt
(+)-neg_115.txt
(+)-neg_116.txt
(-)neg_117.txt
(-)neg_118.txt
(-)neg_119.txt
(-)neg_120.txt
(-)neg_121.txt
(-)neg_122.txt
(-)neg_123.txt
(-)neg_124.txt
(-)neg_125.txt
(-)neg_126.txt
(-)neg_127.txt
(-)neg_128.txt
(-)neg_129.txt
(-)neg_130.txt
(-)neg_131.txt
(-)neg_132.txt
(+)-neg_133.txt
(-)neg_134.txt
(-)neg_135.txt
(-)neg_136.txt
(+)-neg_137.txt
(-)neg_138.txt
(-)neg_139.txt
(-)neg_140.txt
(-)neg_141.txt
(-)neg_142.txt
(+)-neg_143.txt
(-)neg_144.txt
(-)neg_145.txt
(+)-neg_146.txt
(-)neg_147.txt
(-)neg_148.txt
(-)neg_149.txt
(-)neg_150.txt
(+)-neg_151.txt

Text Result | Graphical Result |

Results for selected folder:
This directory is Positive.
Overall accuracy is 87.90%;
Positive precision is 93.00%;
Negative precision is 82.80%.

Number of articles:1000
Number of positive articles:551
Number of negative articles:449
Number of neutral articles:0
Positive percent:55.10%

File: Test

First, I tried to upgrade. There were no seats left, but I still spent over 20 dollars to try and do so, because they're the only flagship airline I know that has a paid (and ridiculously expensive: 1 eur/min) customer service phone line. And it's obviously the only way you can upgrade, no option to do so from their website. The check-in was horrible, with a never ending line because it took them forever to check people in. In the end they had to ask help from the desks of another airline to cope with that. The flight experience was also horrible: the plane was old and filthy, the food was little and disgusting (I very rarely leave food on my plate, but some of that was inedible), the in-flight entertainment consisted of a screen just slightly bigger than my phone with a handful of movies that kept breaking for no apparent reason. The service on the plane was, guess what, horrible. Their stewards were borderline rude, throwing things on my tray when serving them, constantly hitting my head or shoulder because they didn't look where going. They were also arrogant and incompetent: when I asked for an immigration card, they told me I didn't need it anymore, and got annoyed as I insisted. In the end I didn't get one, and once I reached the immigration officer I had to go to the back of the queue because obviously I was right and they were wrong, and I did need an immigration card with my visa. I vowed in the past I wouldn't travel with Alitalia anymore. This time I couldn't avoid it, but they confirmed, once again if there was any need, that I made the right choice back then. I do hope they finally go out of business, as they absolutely deserve that.

[Text Result](#) | [Graphical Result](#)

Overall Document: Test in hybrid-based model result is Negative
Probability to be positive is 0.02% with confidence 99.97%

CabinService: Test in rule-based model result is Negative
Probability to be positive is 16.49% with confidence 79.38%

GroundService: Test in rule-based model result is Negative
Probability to be positive is 11.64% with confidence 85.45%

FoodBeverage: Test in rule-based model result is Negative
Probability to be positive is 30.77% with confidence 61.54%

Entertainment: Test in rule-based model result is Negative
Probability to be positive is 40.00% with confidence 50.00%

24 matches:
6 matches for product definitions:

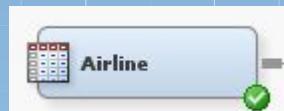
- No 0 (42-46): CabinService : seats
- No 1 (332-339): GroundService : check-in
- No 2 (589-592): FoodBeverage : food

Product prominence information:
GroundService: Top 20%
CabinService: Top 20%
FoodBeverage: Bottom 80%
Entertainment: Bottom 80%

Product dominance information:
GroundService: In Passing
CabinService: In Passing
FoodBeverage: In Passing
Entertainment: In Passing

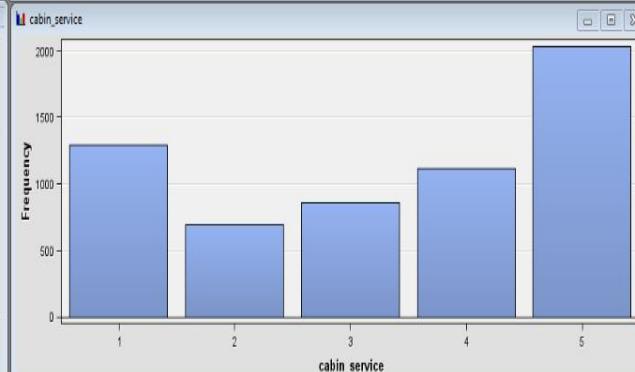
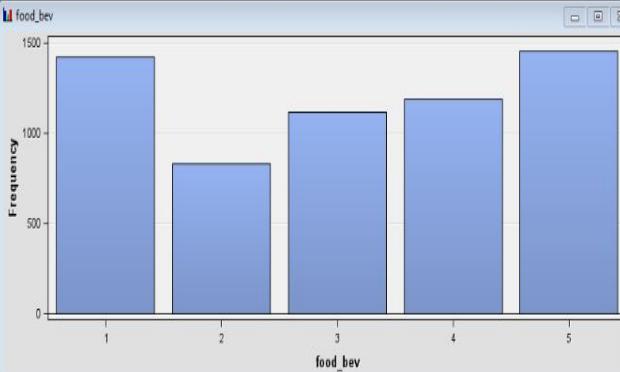
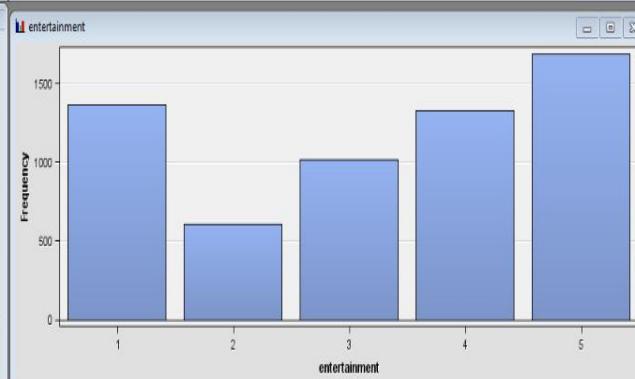
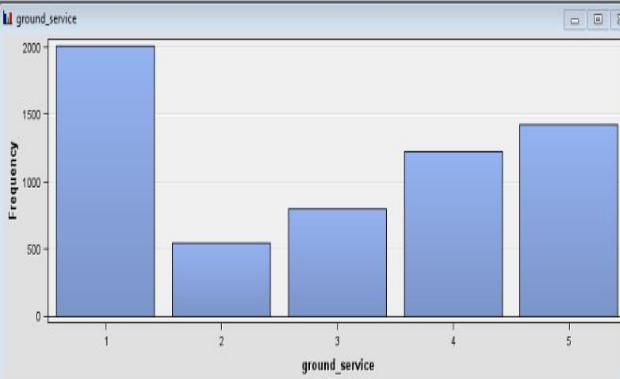
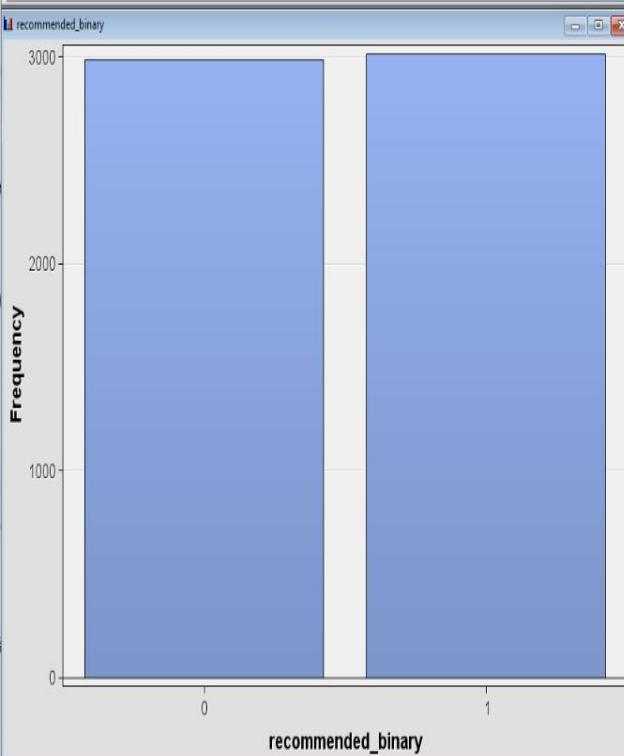
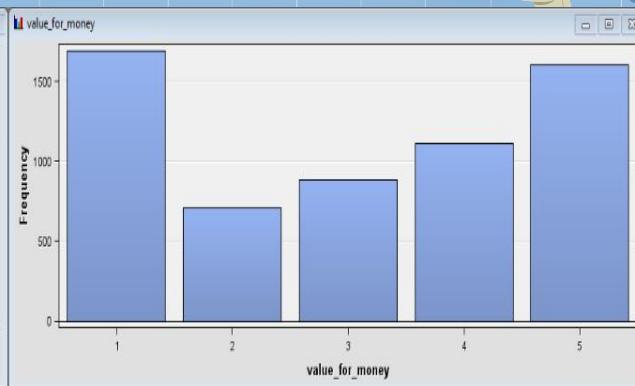
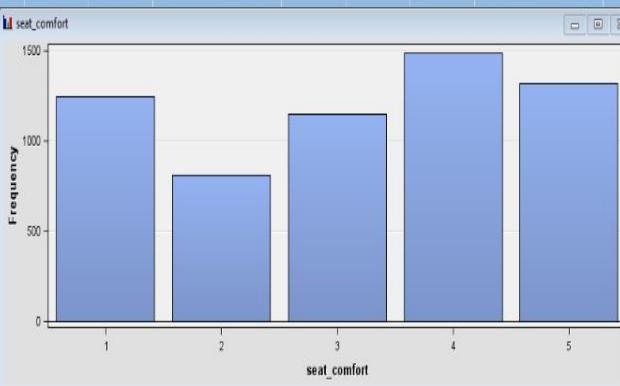
APPENDIX - OBJECTIVE 3

APPENDIX - OBJECTIVE 3



Sample Statistics

Obs #	Variable ...	Label	Type	Percent ...	Minimum	Maximum	Mean
1	cabin_serv...		VAR	0	1	5	3.319
2	entertainm...		VAR	0	1	5	3.229
3	food_bev...		VAR	0	1	5	3.0705
4	ground_ser...		VAR	0	1	5	2.9175
5	recommen...		VAR	0	0	1	0.5025
6	seat_comfort		VAR	0	1	5	3.137
7	value_for_...		VAR	0	1	5	3.038687



APPENDIX - OBJECTIVE 3



Data Set Allocations	
Training	60.0
Validation	20.0
Test	20.0

Output

```
50 Data=DATA
51
52             Numeric   Formatted   Frequency
53     Variable      Value      Value      Count      Percent    Label
54
55 recommended_binary      0          0      11753    51.5279
56 recommended_binary      1          1      11056    48.4721
57
58
59 Data=TEST
60
61             Numeric   Formatted   Frequency
62     Variable      Value      Value      Count      Percent    Label
63
64 recommended_binary      0          0      2350    51.5238
65 recommended_binary      1          1      2211    48.4762
66
67
68 Data=TRAIN
69
70             Numeric   Formatted   Frequency
71     Variable      Value      Value      Count      Percent    Label
72
73 recommended_binary      0          0      7052    51.5271
74 recommended_binary      1          1      6634    48.4729
75 |
76
77 Data=VALIDATE
78
79             Numeric   Formatted   Frequency
80     Variable      Value      Value      Count      Percent    Label
81
82 recommended_binary      0          0      2351    51.5344
83 recommended_binary      1          1      2211    48.4656
84
```

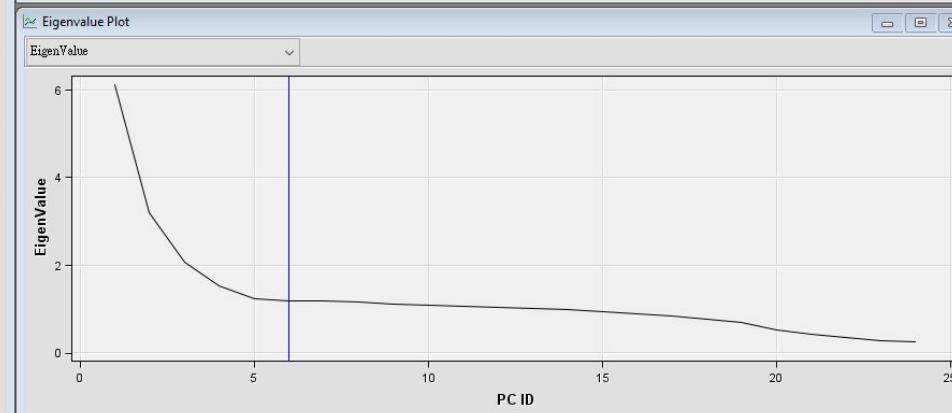
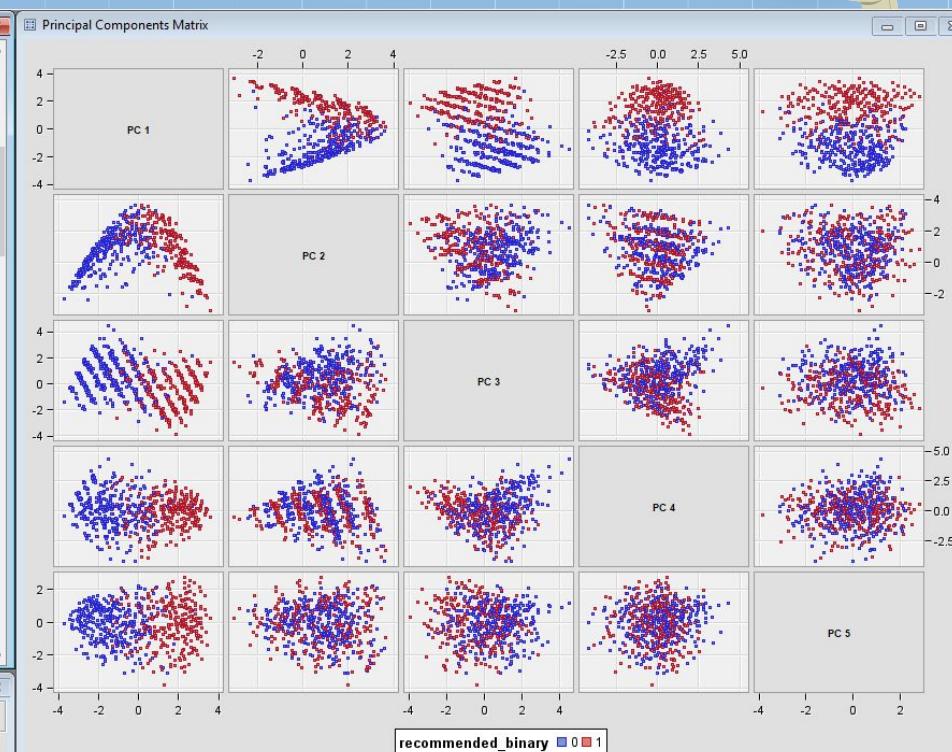
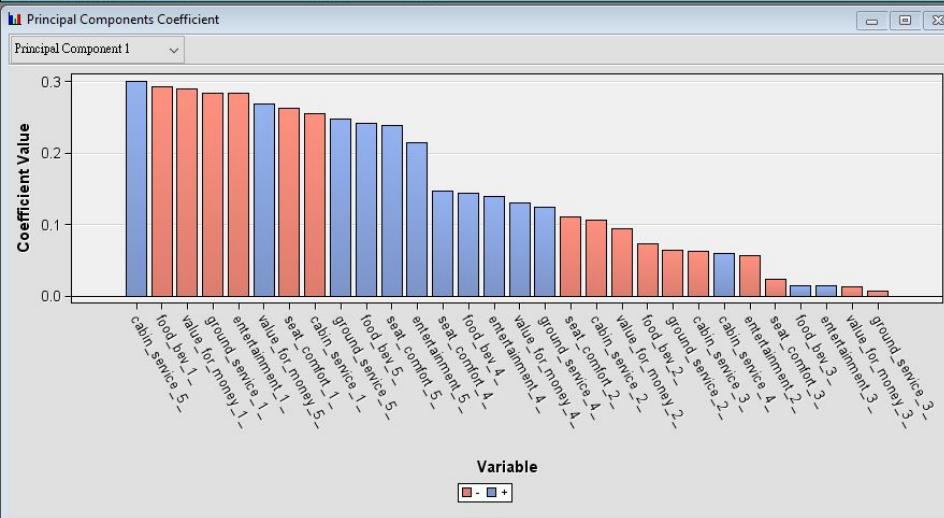
APPENDIX - OBJECTIVE 3



Train	
Variables	Correlation
Eigenvalue Source	Cumulative
Interactive Selection	Increment
Print Eigenvalue Source	No

Eigenvalue Cutoff	0.99
Cumulative	0.99
Increment	0.001

25 The IMNUEUL Procedure
26
27 Variable Level Mean Std Dev
28
29 cabin_service 1 0.22804 0.41959
30 cabin_service 2 0.12305 0.32850
31 cabin_service 3 0.14241 0.34948
32 cabin_service 4 0.17646 0.38122
33 cabin_service 5 0.33005 0.47025
34 entertainment 1 0.29190 0.45466
35 entertainment 2 0.11925 0.32409
36 entertainment 3 0.18150 0.38545
37 entertainment 4 0.21372 0.40995
38 entertainment 5 0.19363 0.39516
39 food_bev 1 0.29804 0.45741
40 food_bev 2 0.13664 0.34348
41 food_bev 3 0.17112 0.37663
42 food_bev 4 0.19173 0.39368
43 food_bev 5 0.20247 0.40186
44 ground_service 1 0.32895 0.46985
45 ground_service 2 0.09535 0.29371
46 ground_service 3 0.14087 0.34790
47 ground_service 4 0.20393 0.40293
48 ground_service 5 0.23089 0.42142
49 seat_comfort 1 0.23615 0.42473
50 seat_comfort 2 0.13707 0.34394
51 seat_comfort 3 0.19056 0.39276
52 seat_comfort 4 0.23250 0.42244
53 seat_comfort 5 0.20271 0.40277
54 value_for_money 1 0.29278 0.45506
55 value_for_money 2 0.12699 0.33298



APPENDIX - OBJECTIVE 3

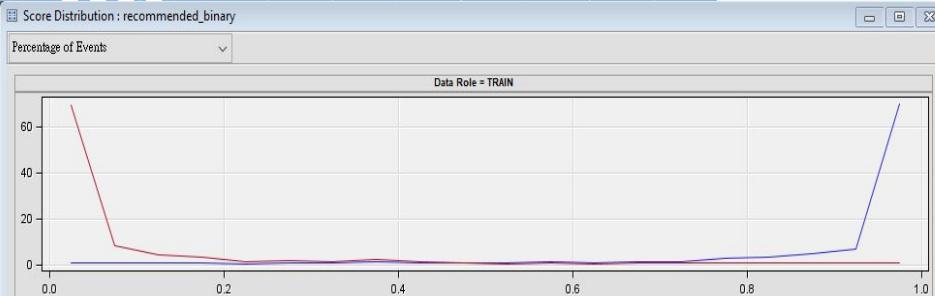


Variable	Label	Principal Component t1 ▼	Principal Component t2	Principal Component t3	Principal Component t4	Principal Component t5	Principal Component t6
cabin_servi... 5		0.301479	-0.20729	0.030971	0.013527	-0.03475	-0.03434
value_for_... 5		0.268562	-0.21866	0.088554	-0.037	-0.06077	0.061649
ground_ser... 5		0.247694	-0.2226	0.113232	-0.03549	-0.15418	-0.01103
food_bev_5_ 5		0.241674	-0.24667	0.145355	-0.05916	0.117351	-0.00361
seat_comfo... 5		0.239081	-0.24182	0.145765	-0.0548	-0.01292	-0.06007
entertainm... 5		0.215295	-0.20392	0.132605	-0.06395	0.219499	0.076649
seat_comfo... 4		0.14739	0.142087	-0.29634	0.16362	0.078834	0.121586
food_bev_4_ 4		0.144295	0.111707	-0.30599	0.224603	-0.20897	0.001512
entertainm... 4		0.139415	0.098673	-0.2231	0.202959	-0.44832	-0.18779
value_for_... 4		0.129762	0.168681	-0.27484	0.086444	0.146256	-0.10464
ground_ser... 4		0.124314	0.159057	-0.25438	0.066229	0.427918	-0.06425
cabin_servi... 4		0.05883	0.232734	-0.23714	0.036153	0.121265	0.084549
food_bev_3_ 3		0.013778	0.234626	0.023712	-0.41809	0.124682	0.056593
entertainm... 3		0.01363	0.200712	0.011267	-0.32224	0.228243	0.301388
ground_ser... 3		-0.00625	0.177524	0.071132	-0.08378	-0.47963	0.092217
value_for_... 3		-0.0122	0.220518	0.117643	-0.15098	-0.13504	-0.01513
seat_comfo... 3		-0.02331	0.234195	0.097726	-0.32186	-0.14806	-0.12882
entertainm... 2		-0.05713	0.129704	0.215843	0.158507	0.155955	-0.40551
cabin_servi... 3		-0.06198	0.191563	0.170942	-0.2886	-0.13839	-0.08855
ground_ser... 2		-0.06403	0.09898	0.196805	0.220487	0.035757	0.43728
food_bev_2_ 2		-0.07354	0.159327	0.286406	0.22361	0.023878	-0.31265
value_for_... 2		-0.09467	0.113735	0.240029	0.226708	-0.04353	0.373041
cabin_servi... 2		-0.10665	0.080883	0.24036	0.28624	0.068381	-0.03852
seat_comfo... 2		-0.11008	0.093861	0.268705	0.281421	0.146749	-0.00103
cabin_servi... 1		-0.2562	-0.20202	-0.14982	-0.03173	-0.0095	0.065581
seat_comfo... 1		-0.26262	-0.20458	-0.15145	-0.04103	-0.04807	0.055995
entertainm... 1		-0.28366	-0.17435	-0.0775	0.032775	-0.09121	0.136258
ground_ser... 1		-0.28412	-0.13007	-0.05911	-0.10077	0.104108	-0.27665
value_for_... 1		-0.29048	-0.18489	-0.11291	-0.09093	0.066146	-0.23069
food_bev_1_ 1		-0.29263	-0.19226	-0.09894	0.035013	-0.04384	0.190042

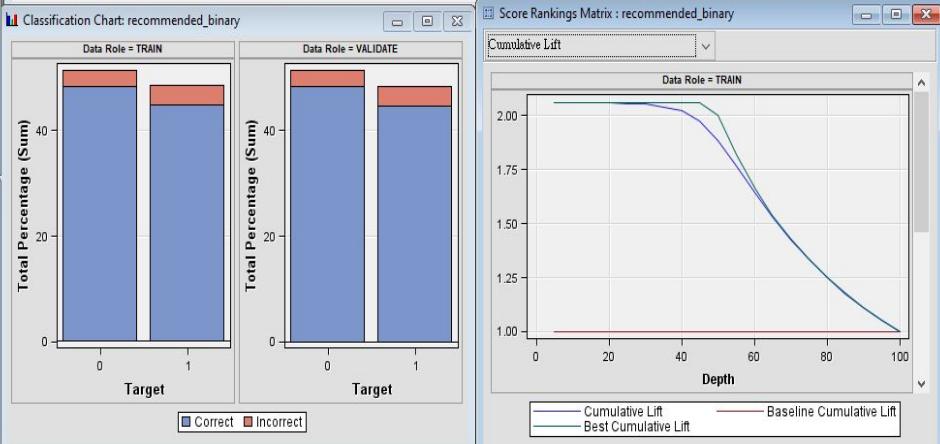
APPENDIX - OBJECTIVE 3



Method	Scan
Number of Neighbors	150

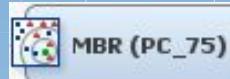


Target	Fit Statistics	Statistics Label	Train	Validation	Test	Target Label
recommended_binary _NW_	Number of Estimated ...		6			
recommended_binary _NQBS_	Sum of Frequencies		13686	4562	4561	
recommended_binary _SUMW_	Sum of Case Weights ...		27372	9124	9122	
recommended_binary _DFT_	Total Degrees of Free...		13686			
recommended_binary _DFM_	Model Degrees of Fre...		6			
recommended_binary _DFE_	Degrees of Freedom f...		13680			
recommended_binary _ASE_	Average Squared Error		0.050303	0.052242	0.050066	
recommended_binary _RASE_	Root Average Squared...		0.224283	0.228564	0.223755	
recommended_binary _DIV_	Divisor for ASE		27372	9124	9122	
recommended_binary _SSE_	Sum of Squared Errors		1376.89	476.6524	456.7028	
recommended_binary _MSE_	Mean Squared Error		0.050325	0.052242	0.050066	
recommended_binary _RMSE_	Root Mean Squared Er...		0.224332	0.228564	0.223755	
recommended_binary _AVER_	Average Error Function		0.167994	0.17867	0.169098	
recommended_binary _ERR_	Error Function		4598.343	1630.181	1542.515	
recommended_binary _MAX_	Maximum Absolute Err...		1	1	1	
recommended_binary _FPE_	Final Prediction Error		0.050347			
recommended_binary _RFPE_	Root Final Prediction ...		0.224381			
recommended_binary _AIC_	Akaike's Information C...		4610.343			
recommended_binary _SBC_	Schwarz's Bayesian C...		4655.488			
recommended_binary _MISC_	Misclassification Rate		0.067441	0.069487	0.066652	
recommended_binary _WRONG_	Number of Wrong Cla...		923	317	304	

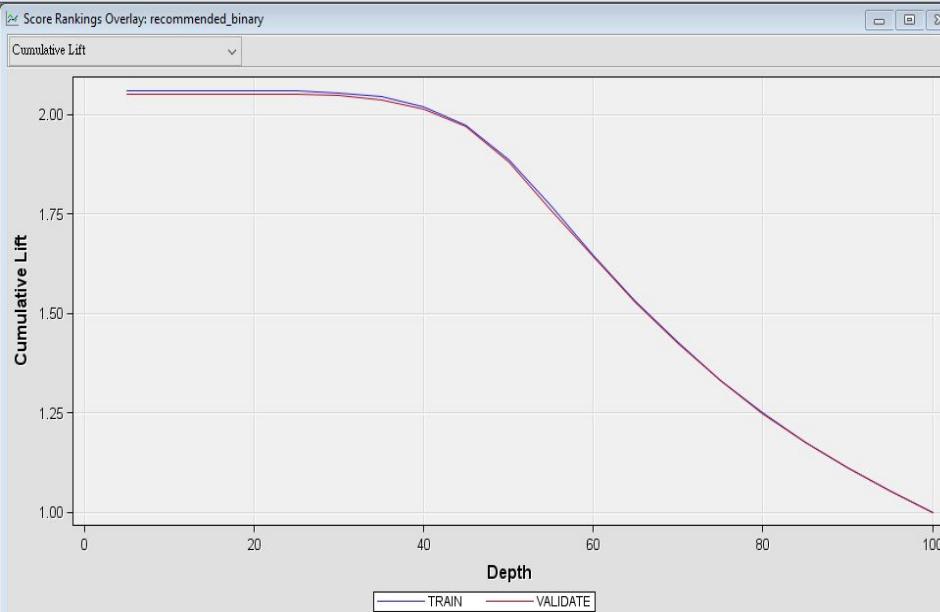
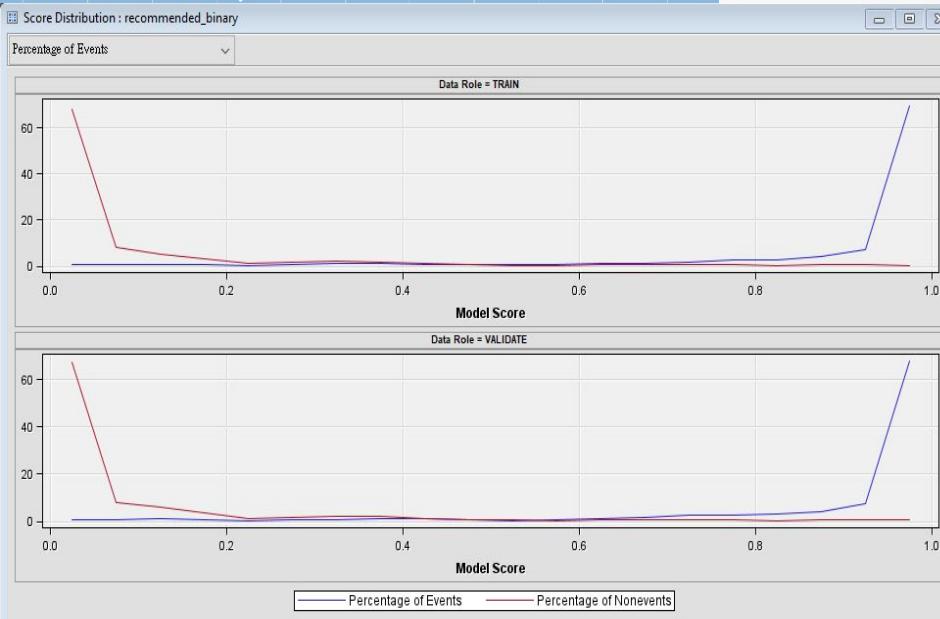


91	Classification Table						
92	Data Role=TRAIN Target Variable=recommended_binary Target Label='1'						
93	Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage	
94	99	0	93.0710	93.9024	6622	48.3852	
95	100	1	6.9290	7.4314	493	3.6022	
96	101	0	6.5439	6.0976	430	3.1419	
97	102	1	93.4561	92.5686	6141	44.8707	
98							
99	Data Role=VALIDATE Target Variable=recommended_binary Target Label='1'						
100							
101							
102							
103							
104							
105							

APPENDIX - OBJECTIVE 3

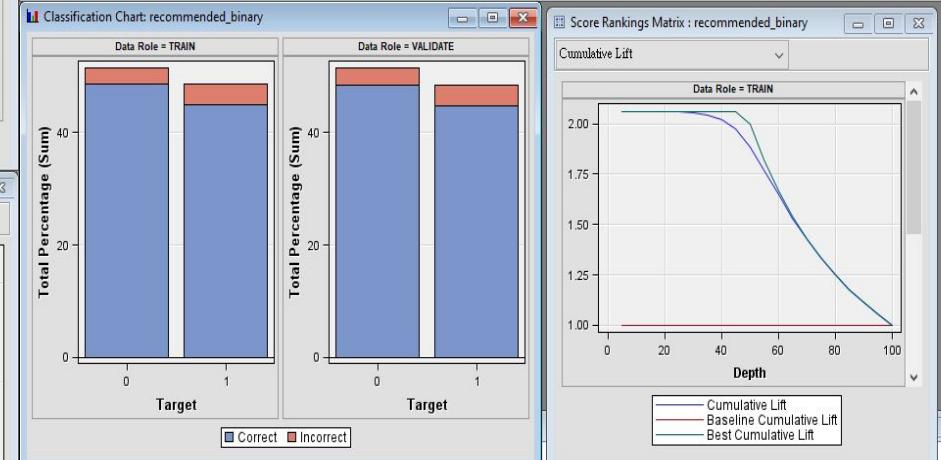


Method	Scan
Number of Neighbors	75



Fit Statistics

Target	Fit Statistics	Statistics Label	Train	Validation	Test	Target Label
recommended_binary	_NW_	Number of Estimated ...	6			
recommended_binary	_NOBS_	Sum of Frequencies ...	13686	4562	4561	
recommended_binary	_SUMWV_	Sum of Case Weights ...	27372	9124	9122	
recommended_binary	_DFT_	Total Degrees of Free...	13686			
recommended_binary	_DFM_	Model Degrees of Fre...	6			
recommended_binary	_DFE_	Degrees of Freedom f...	13680			
recommended_binary	_ASE_	Average Squared Error	0.049598	0.052121	0.050349	
recommended_binary	_RASE_	Root Average Squared...	0.222706	0.2283	0.224386	
recommended_binary	_DIV_	Divisor for ASE	27372	9124	9122	
recommended_binary	_SSE_	Sum of Squared Errors	1357.593	475.5509	459.2843	
recommended_binary	_MSE_	Mean Squared Error	0.04962	0.052121	0.050349	
recommended_binary	_RMSE_	Root Mean Squared Er...	0.222755	0.2283	0.224386	
recommended_binary	_AVERR_	Average Error Function	0.164244	0.179252	0.170476	
recommended_binary	_ERR_	Error Function	4495.697	1635.491	1555.084	
recommended_binary	_MAX_	Maximum Absolute Err...	1	1	1	
recommended_binary	_FPE_	Final Prediction Error	0.049641			
recommended_binary	_RFPE_	Root Final Prediction ...	0.222803			
recommended_binary	_AIC_	Akaike's Information C...	4507.697			
recommended_binary	_SBC_	Schwarz's Bayesian C...	4552.842			
recommended_binary	_MISC_	Misclassification Rate	0.066711	0.069487	0.067091	
recommended_binary	_WRONG_	Number of Wrong Cla...	913	317	306	



Cumulative Lift

Data Role = TRAIN

Depth

Cumulative Lift — Baseline Cumulative Lift — Best Cumulative Lift

Classification Table

Data Role=TRAIN Target Variable=recommended_binary Target Label='0'

Target	Outcome	Target	Outcome	Frequency	Total
Target	Outcome	Percentage	Percentage	Count	Percentage
99	0	93.0565	94.0726	6634	48.4729
100	1	6.9435	7.4616	495	3.6168
101	0	6.3749	5.9274	418	3.0542
102	1	93.6251	92.5384	6139	44.8561

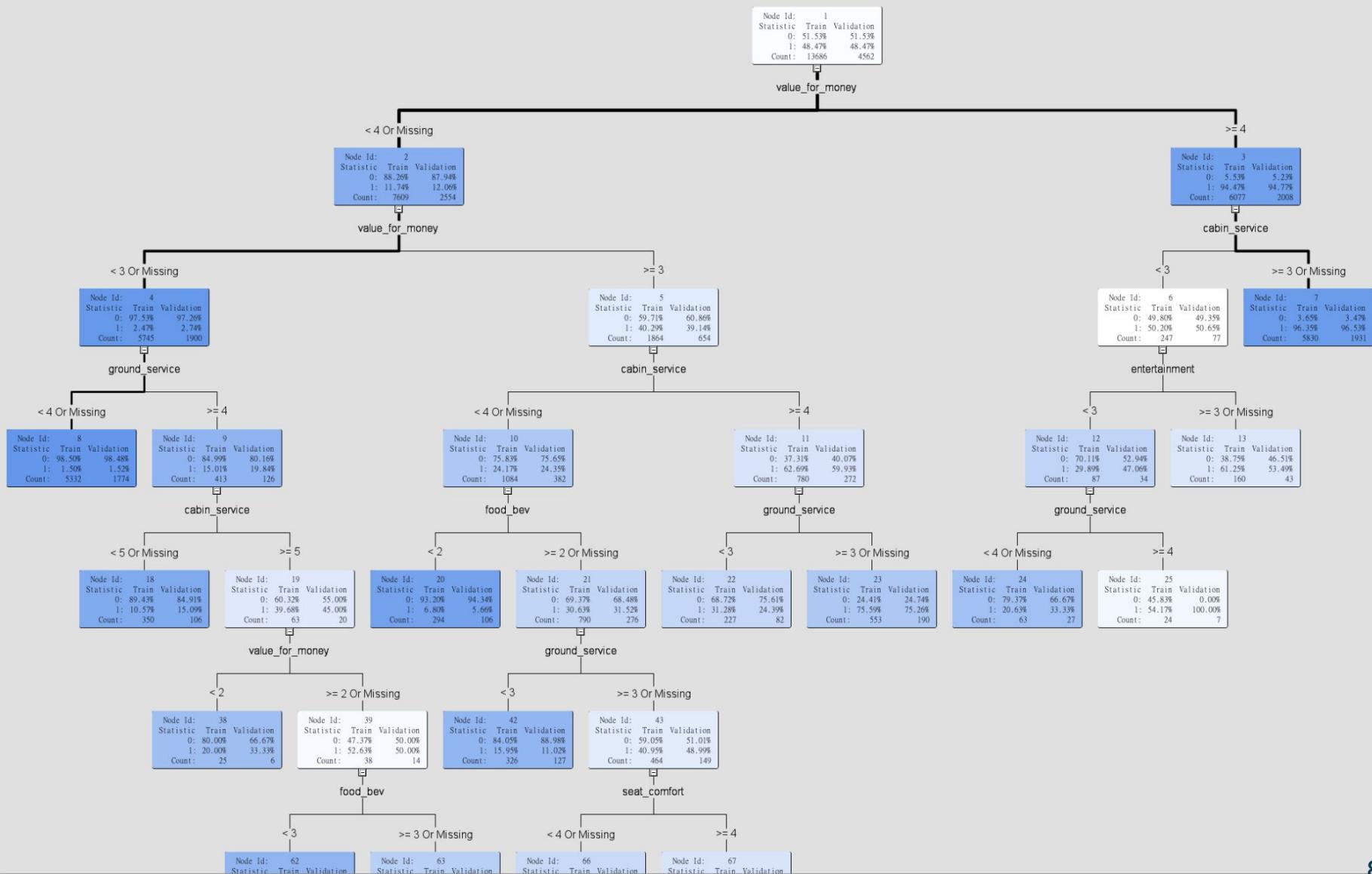
Data Role=VALIDATE Target Variable=recommended_binary Target Label='1'

APPENDIX - OBJECTIVE 3

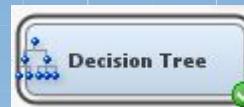


Subtree	
Method	N
Number of Leaves	15
Assessment Measure	Misclassification
Assessment Fraction	0.25

Tree



APPENDIX - OBJECTIVE 3



Score Rankings Overlay: recommended_binary

Cumulative Lift

Depth	Cumulative Lift (TRAIN)	Cumulative Lift (VALIDATE)
0	2.0	2.0
20	2.0	2.0
40	2.0	2.0
60	1.8	1.8
80	1.2	1.2
100	1.0	1.0

Variable Importance

Variable Name	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation Importance	Label
value_for_money	3	1.0000	1.0000	1.0000	
cabin_service	3	0.2205	0.2110	0.9570	
ground_service	4	0.1445	0.1849	1.2800	
food_bev	2	0.0770	0.0831	1.0799	
seat_comfort	1	0.0582	0.0428	0.7352	
entertainment	1	0.0469	0.0000	0.0000	

Fit Statistics

Target	Fit Statistics	Statistics Label	Train	Validation	Test	Target Label
recommended_binary	_NOBS_	Sum of Frequencies	13686	4562	4561	
recommended_binary	_MISC_	Misclassification Rate	0.063203	0.064007	0.064898	
recommended_binary	_MAX_	Maximum Absolute Error	0.984996	0.984996	0.984996	
recommended_binary	_SSE_	Sum of Squared Errors	1397.889	468.4199	468.9305	
recommended_binary	_ASE_	Average Square Error	0.05107	0.051339	0.051407	
recommended_binary	_RASE_	Root Average Squared Error	0.225987	0.226582	0.22673	
recommended_binary	_DIV_	Divisor for ASE	27372	9124	9122	
recommended_binary	_DFT_	Total Degrees of Freedom	13686	-	-	

Output

Tree Leaf Report

Node	Depth	Training Observations	Percent	Validation Observations	Percent
7	2	5810	0.96	1931	0.97
8	3	5332	0.02	1774	0.02
23	4	553	0.76	190	0.75
18	4	350	0.11	106	0.15
66	6	335	0.33	102	0.42
42	5	326	0.16	127	0.11
20	4	294	0.07	106	0.06
22	4	227	0.31	82	0.34
13	3	160	0.61	43	0.53
67	6	129	0.63	47	0.64
24	4	63	0.21	27	0.33
38	5	35	0.20	6	0.33
63	6	25	0.72	9	0.67
25	4	24	0.54	7	1.00
62	6	13	0.15	5	0.30

Leaf Statistics

Index	Training Percent 1	Validation Percent 1
1	0.95	0.95
2	0.75	0.75
3	0.72	0.68
4	0.63	0.65
5	0.60	0.53
6	0.55	1.00
7	0.35	0.45
8	0.30	0.25
9	0.22	0.35
10	0.20	0.35
11	0.15	0.12
12	0.15	0.20
13	0.12	0.18
14	0.08	0.08
15	0.05	0.05

Treemap

Fit Statistics

Target	Fit Statistics	Statistics Label	Train	Validation	Test	
recommended_binary	_NOBS_	Sum of Frequencies	13686.00	4562.00	4561.00	
recommended_binary	_MISC_	Misclassification Rate	0.06	0.06	0.06	
recommended_binary	_MAX_	Maximum Absolute Error	0.98	0.98	0.98	
recommended_binary	_SSE_	Sum of Squared Errors	1397.889	468.42	468.93	
recommended_binary	_ASE_	Average Squared Error	0.05	0.05	0.05	
recommended_binary	_RASE_	Root Average Squared Error	0.23	0.23	0.23	
recommended_binary	_DIV_	Divisor for ASE	27372.00	9124.00	9122.00	
recommended_binary	_DFT_	Total Degrees of Freedom	13686.00	-	-	

Classification Table

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	94.4149	93.2501	6576	48.0491
1	0	5.5851	5.0637	389	2.0423

APPENDIX - OBJECTIVE 3



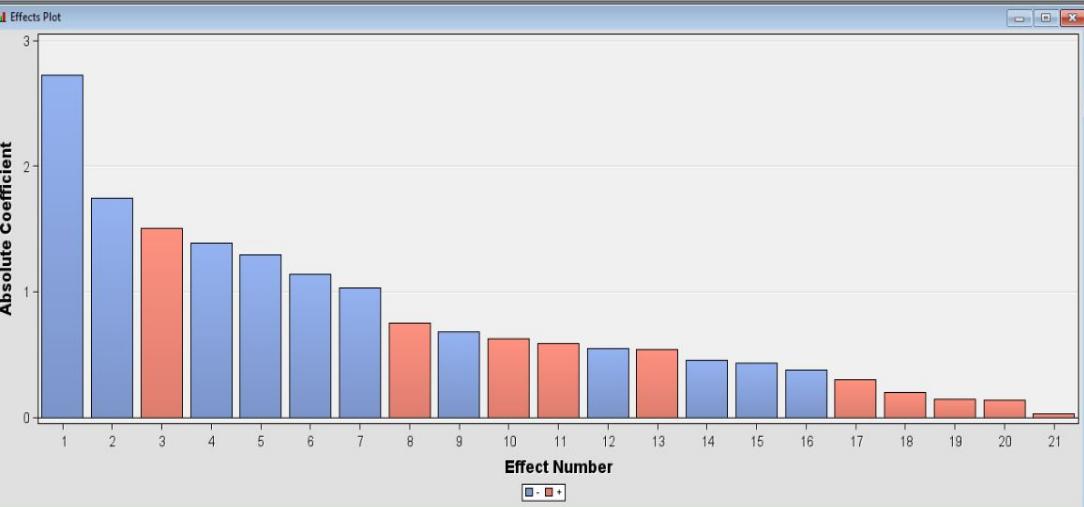
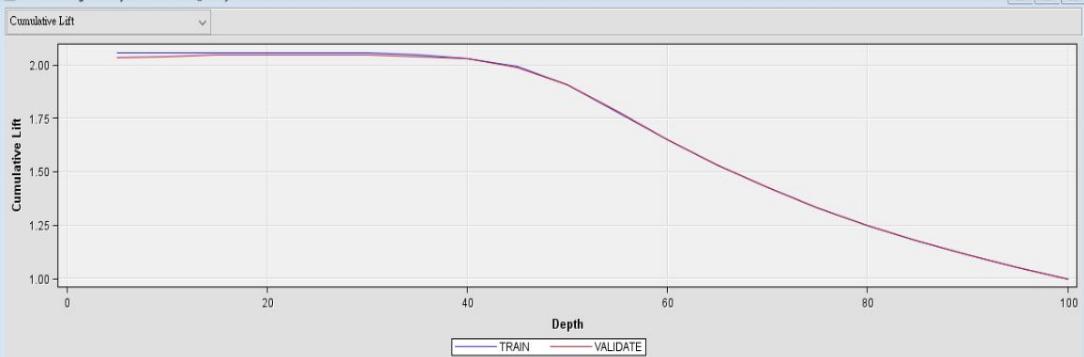
Class Targets
Regression Type
Link Function

Logistic Regression
Logit

Model Selection
Selection Model
Selection Criterion

Stepwise
Validation Misclassification

Score Rankings Overlay: recommended_binary



Fit Statistics

Target	Fit Statistics	Statistics Label	Train	Validation	Test	Target Label
recommended_binary	_AIC_	Akaike's Information Criterion	4146.599			
recommended_binary	_ASE_	Average Squared Error	0.043603	0.043711	0.042487	
recommended_binary	_AVERR_	Average Error Function	0.149956	0.154831	0.145293	
recommended_binary	_DFE_	Degrees of Freedom for Error	13665			
recommended_binary	_DFM_	Model Degrees of Freedom	21			
recommended_binary	_DFT_	Total Degrees of Freedom	13686			
recommended_binary	_DIV_	Divisor for ASE	27372	9124	9122	
recommended_binary	_ERR_	Error Function	4104.599	1412.677	1325.359	
recommended_binary	_FPE_	Final Prediction Error	0.043737			
recommended_binary	_MAX_	Maximum Absolute Error	0.099818	0.099818	0.097355	
recommended_binary	_MSE_	Mean Square Error	0.04367	0.043711	0.042487	
recommended_binary	_NOBS_	Sum of Frequencies	13686	4562	4561	
recommended_binary	_NW_	Number of Estimate Weights	21			
recommended_binary	_RASE_	Root Average Sum of Squares	0.208814	0.209071	0.206124	
recommended_binary	_RFPE_	Root Final Prediction Error	0.209135			
recommended_binary	_RMSE_	Root Mean Squared Error	0.208974	0.209071	0.206124	
recommended_binary	_SBC_	Schwarz's Bayesian Criterion	4304.606			
recommended_binary	_SSE_	Sum of Squared Errors	1193.51	398.8154	387.5657	
recommended_binary	_SUMW_	Sum of Case Weights Times F...	27372	9124	9122	
recommended_binary	_MISC_	Misclassification Rate	0.059112	0.057431	0.057882	

Output

Summary of Stepwise Selection

Step	Entered	IF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Validation Misclassification Rate
1	value_for_money	4	1	10132.7723	<.0001	<.0001	0.0905
2	cabin_service	4	2	1003.2694	<.0001	<.0001	0.0776
3	ground_service	4	3	580.8184	<.0001	<.0001	0.0640
4	food_bev	4	4	239.6454	<.0001	<.0001	0.0598
5	seat_comfort	4	5	156.2927	<.0001	<.0001	0.0574
6	entertainment	4	6	29.7509	<.0001	<.0001	0.0577

The selected model, based on the misclassification rate for the validation data, is the model trained in Step 5. It consists of the following effects:

Intercept cabin_service food_bev ground_service seat_comfort value_for_money

Likelihood Ratio Test for Global Null Hypothesis: $\text{BETA}=0$

Effect	IF	Chi-Square	Pr > ChiSq
Intercept	Only	18960.056	<.0001
cabin_service	Covariates	4104.599	<.0001
food_bev		14855.4571	<.0001

Type 3 Analysis of Effects

Effect	IF	Chi-Square	Pr > ChiSq
cabin_service	4	221.0754	<.0001
food_bev	4	166.5837	<.0001
ground_service	4	456.7955	<.0001
seat_comfort	4	148.7308	<.0001
value_for_money	4	890.4783	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	IF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	-0.6776	0.0697	94.44	<.0001	0.508
cabin_service	1	-1.2911	0.1317	96.11	<.0001	0.275
cabin_service	2	-0.5495	0.0990	30.78	<.0001	0.577
cabin_service	3	0.1474	0.0791	3.47	0.0625	1.159
cabin_service	4	0.5060	0.0760	59.61	<.0001	1.798
food_bev	1	-1.0333	0.1020	102.66	<.0001	0.356
food_bev	2	0.4573	0.0846	29.21	<.0001	0.633
food_bev	3	0.1323	0.0742	3.18	0.0744	1.141
food_bev	4	0.6260	0.0838	55.86	<.0001	1.870
ground_service	1	-1.7444	0.0945	340.59	<.0001	0.175
ground_service	2	-0.3773	0.0952	15.72	<.0001	0.696
ground_service	3	0.2971	0.0749	15.72	<.0001	1.346
ground_service	4	0.7496	0.0744	101.42	<.0001	2.116
seat_comfort	1	-1.1395	0.1262	81.51	<.0001	0.320
seat_comfort	2	-0.4305	0.0961	20.10	<.0001	0.650
seat_comfort	3	0.2000	0.0742	7.26	0.0070	1.221
seat_comfort	4	0.5387	0.0771	48.86	<.0001	1.714
value_for_money	1	-2.7270	0.1688	260.95	<.0001	0.065
value_for_money	2	-1.3061	0.1054	178.85	<.0001	0.250
value_for_money	3	0.0270	0.0767	0.12	0.7252	1.027
value_for_money	4	1.5025	0.0845	316.11	<.0001	4.493

APPENDIX - OBJECTIVE 3



General	
Node ID	HPDMForest3
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Tree Options	
Maximum Number of Trees	21
Seed	12345
Type of Sample	Proportion
Proportion of Obs in Each Samp	0.4
Number of Obs in Each Samp	
Splitting Rule Options	
Maximum Depth	50
Missing Values	Use In Search
Minimum Use In Search	1
Number of Variables to Consider	2
Significance Level	0.05
Max Categories in Split Search	30
Minimum Category Size	5
Exhaustive	5000
Node Options	
Method for Leaf Size	Default
Smallest Percentage of Obs in Node	1.0E-5
Smallest Number of Obs in Node	
Split Size	.
Use as Modeling Node	Yes
Score	
Variable Selection	Yes
Variable Importance Method	Loss Reduction
Number of Variables to Consider	25
Cutoff Fraction	0.01

APPENDIX - OBJECTIVE 3



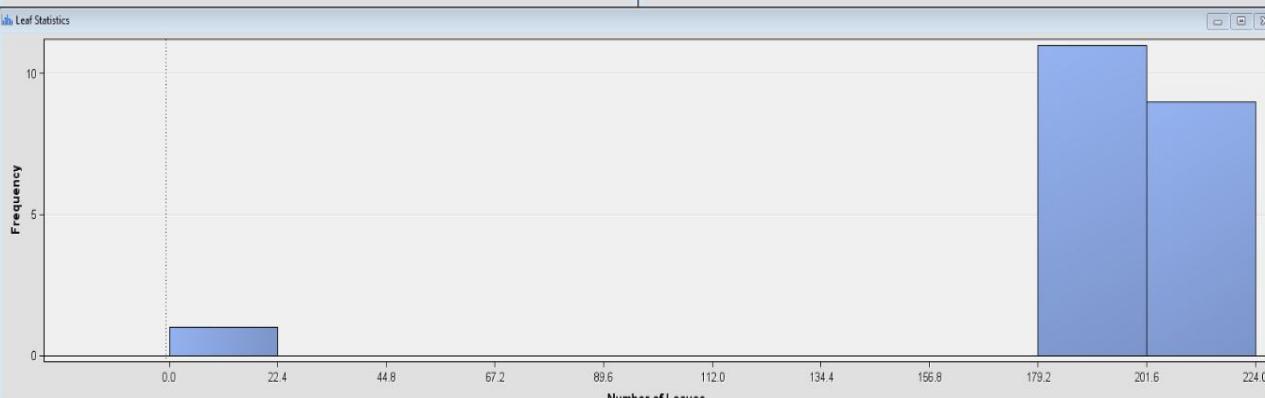
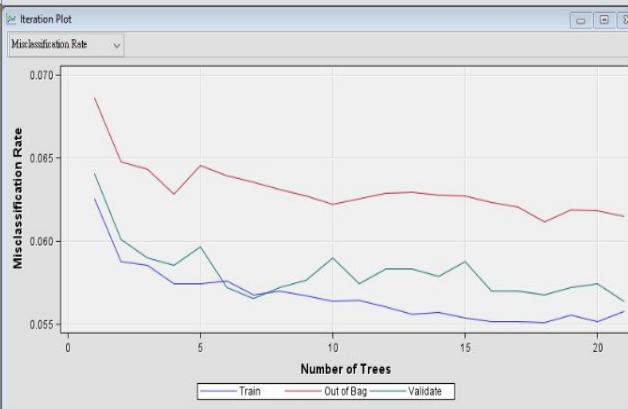
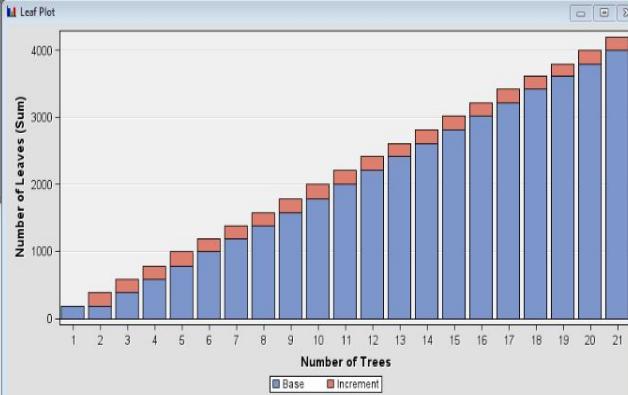
Number of Trees	Number of Leaves	Average Error (Train)	Average Error (Out of Bag)	Misclassification Rate (Train)	Misclassification Rate (Out of Bag)	Log Loss (Train)	Log Loss (Out of Bag)	Log Loss (Validate)
21	4199	0.0404	0.0455	0.0445	0.0558	0.0614	0.0563	0.135
7	1385	0.0414	0.0469	0.0456	0.0568	0.0635	0.0566	0.137
18	3814	0.0404	0.0455	0.0445	0.0551	0.0612	0.0568	0.135
16	3208	0.0405	0.0457	0.0445	0.0552	0.0623	0.0570	0.136
17	3421	0.0404	0.0456	0.0444	0.0552	0.0620	0.0570	0.135
6	1188	0.0415	0.0472	0.0458	0.0576	0.0639	0.0572	0.138
8	1583	0.0413	0.0468	0.0453	0.0570	0.0631	0.0572	0.137
19	3799	0.0404	0.0455	0.0445	0.0555	0.0619	0.0572	0.135
11	2208	0.0409	0.0461	0.0449	0.0564	0.0624	0.0574	0.134
20	3993	0.0404	0.0455	0.0444	0.0552	0.0618	0.0574	0.135
9	1780	0.0411	0.0465	0.0452	0.0567	0.0627	0.0577	0.137
14	2811	0.0406	0.0459	0.0448	0.0557	0.0628	0.0579	0.136
12	2413	0.0403	0.0461	0.0449	0.0560	0.0628	0.0583	0.136
13	2603	0.0407	0.0460	0.0449	0.0556	0.0629	0.0583	0.136
4	784	0.0418	0.0478	0.0464	0.0574	0.0628	0.0585	0.141
15	3024	0.0405	0.0457	0.0446	0.0554	0.0627	0.0587	0.135
3	583	0.0423	0.0469	0.0467	0.0585	0.0643	0.0590	0.151
10	2004	0.0410	0.0463	0.0449	0.0563	0.0622	0.0590	0.137
5	997	0.0417	0.0480	0.0462	0.0574	0.0645	0.0566	0.139
						0.139	0.218	0.190

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
recommended_bin...	_ASE_	Average Squared E...		0.040372	0.044463	0.043103
recommended_bin...	_DIV_	Divisor for ASE		27372	9124	9122
recommended_bin...	_MAX_	Maximum Absolute...		0.99828	0.999932	0.999159
recommended_bin...	_NOBS_	Sum of Frequencies		13886	4562	4561
recommended_bin...	_RASE_	Root Average Squa...		0.20927	0.210863	0.207613
recommended_bin...	_SSE_	Sum of Squared Err...		1105.056	405.6808	393.1864
recommended_bin...	_DISF_	Frequency of Class...		13886	4562	4561
recommended_bin...	_MISC_	Misclassification R...		0.05575	0.056335	0.059198
recommended_bin...	_WRONG_	Number of Wrong ...		763	257	270

Variable Importance

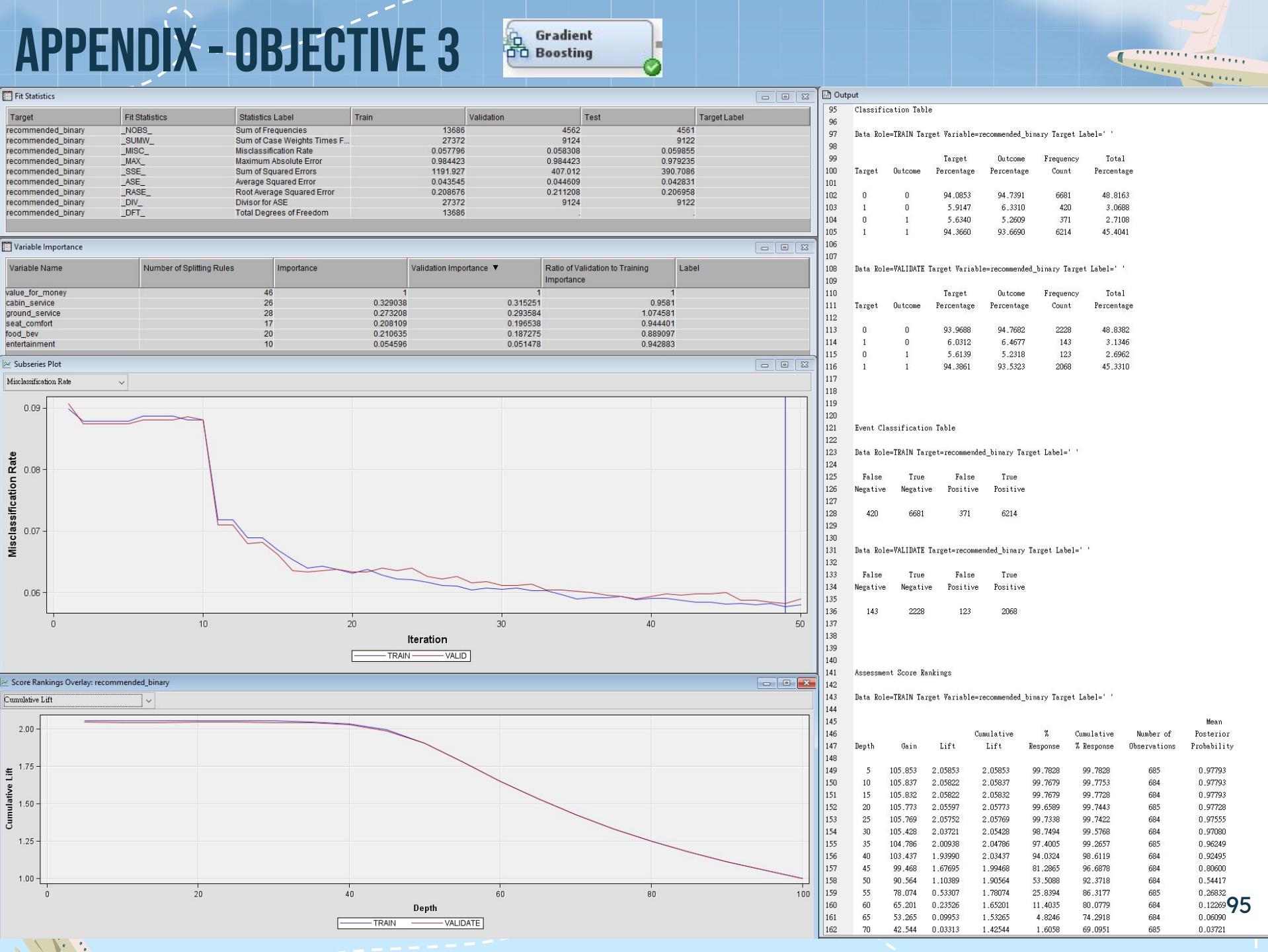
Variable Name	Number of Splitting Rules	Train: Gini Reduction	Train: Margin Reduction	OOB: Gini Reduction	OOB Margin Reduction	Valid: Gini Reduction	Valid: Margin Reduction	Label
value_for_mon...	620	0.163905	0.327811	0.15797	0.322589	0.15877	0.322213	
cabin_sevice	690	0.084388	0.168775	0.08115	0.165750	0.07944	0.164654	
ground_service	698	0.072204	0.144408	0.06757	0.139235	0.07481	0.148369	
seat_comfort	679	0.061302	0.122783	0.05516	0.116369	0.05368	0.113887	
food_bev	683	0.031542	0.063084	0.02939	0.060705	0.02598	0.057599	
entertainment	698	0.065536	0.011072	-0.00050	0.005641	-0.00095	0.004330	



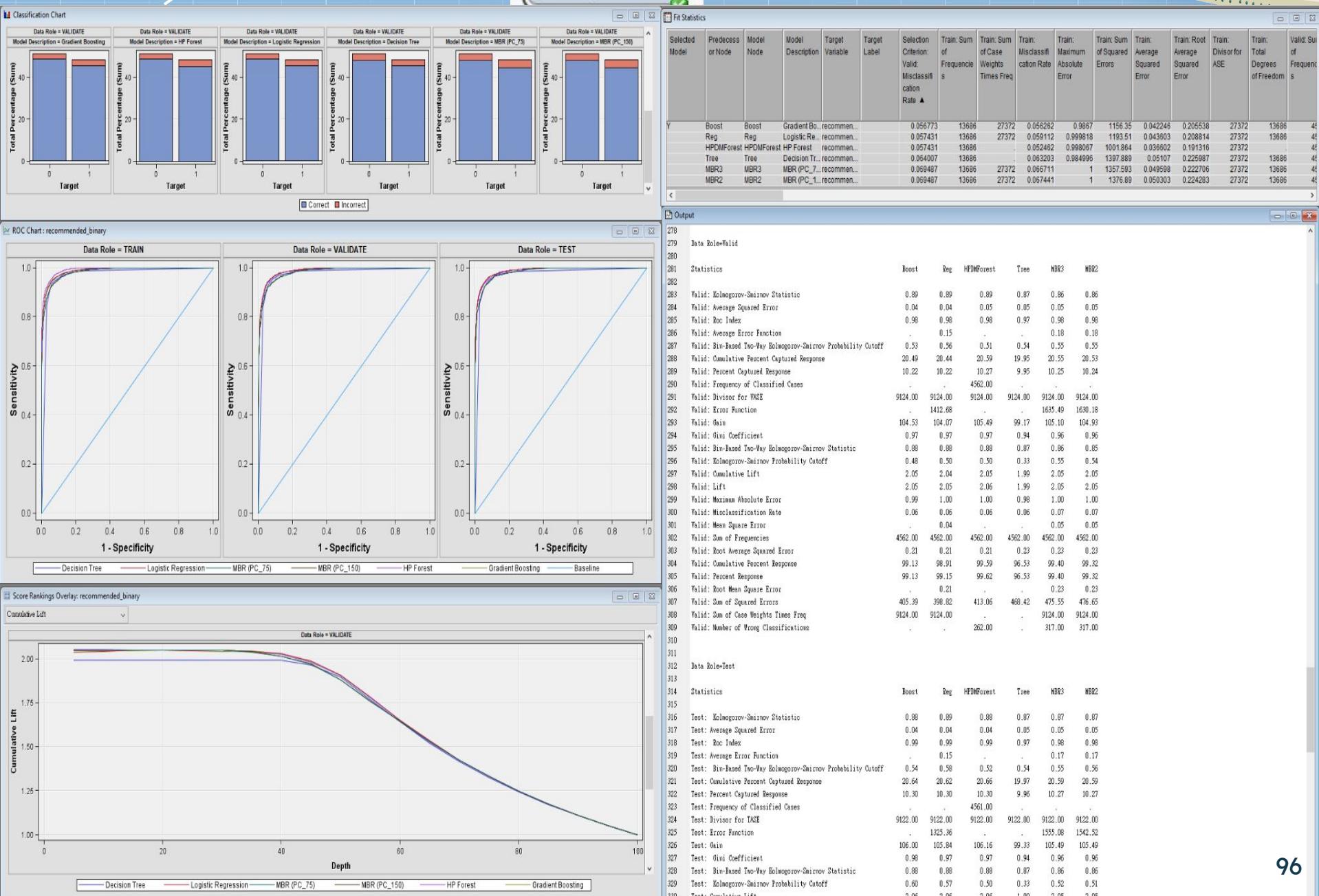
APPENDIX - OBJECTIVE 3



Train	
Variables	[...]
Series Options	
N Iterations	49
Seed	12345
Shrinkage	0.1
Train Proportion	100
Splitting Rule	
Huber M-Regression	No
Maximum Branch	2
Maximum Depth	3
Minimum Categorical Size	5
Reuse Variable	1
Categorical Bins	30
Interval Bins	100
Missing Values	Use in search
Performance	Disk
Node	
Leaf Fraction	0.001
Number of Surrogate Rules	2
Split Size	100
Split Search	
Exhaustive	5000
Node Sample	20000
Subtree	
Assessment Measure	Decision
Score	
Subseries	Best Assessment Value
Number of Iterations	1
Create H Statistic	No
Variable Selection	Yes



APPENDIX - OBJECTIVE 3



VARIABLE IMPORTANCE

	Gradient Boosting (Validation Importance)	HP Forest (Valid: Gini Reduction)	Decision Tree (Validation Importance)
Variable name	Value_for_money	Value_for_money	Value_for_money
Score	1	0.15877	1
	Seat_comfort	ground_service	Cabin_service
Score	0.89882	0.07944	0.21101
	Cabin_service	Cabin_service	Ground_service
Score	0.89204	0.07481	0.18495
	Ground_service	Seat_comfort	Food_bev
Score	0.36253	0.05386	0.08314
	Food_bev	Food_bev	Seat_comfort
Score	0.29605	0.02598	0.04280
	Entertainment	Entertainment	Entertainment
Score	0.08254	-0.00098	0

APPENDIX - OBJECTIVE 4

JARGON EXPLANATION

Aircraft maintenance

Amm: aircraft maintenance manual

Srm: structural repair manual

Doubler: A piece of sheet metal used to strengthen and stiffen a repair in a sheet metal structure

Iaw: maintenance guidance

Flying a plane

OPS: Rules for designing instrument approach and departure procedures

Plane component

APU: Auxiliary Power Unit Dwg: device working group

Type of cargo

Atf cargo Fwd cargo

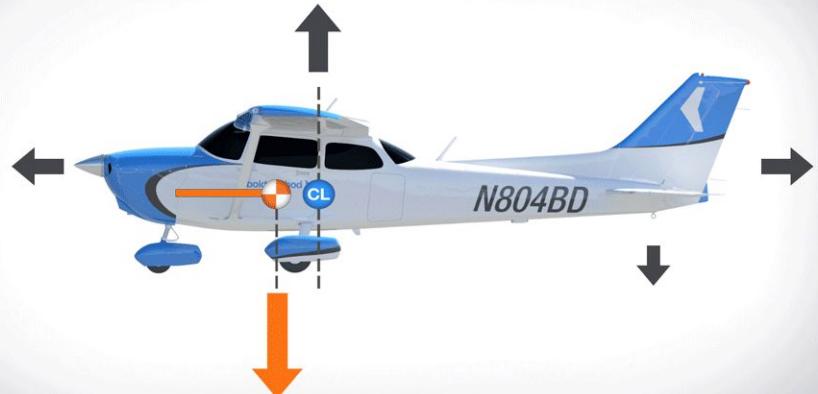
Number+r = runway



JARGON EXPLANATION

AFT:

Aft CG



boldmethod ➤



JARGON EXPLANATION

Crew

INOP: refers to an item of aircraft equipment or a system that is partially or totally unavailable to the pilot or crew.

FCB: Flight Crew Bulletin

XOS: aircraft maintenance provider

RII: Required Inspection Item

Other

BR: Basic regulations

WI: whitelist STR: Standardized Taxi Routes

STA: Scheduled Time of Arrival

BTW: Batu Licin Airport

Rt: Radio transmit



APPENDIX - OBJECTIVE 4

Stop list:

1. Airplane name
2. Airport name
3. Runway number
4. Repair manual short form
5. Section no. in the repair manual

Total: 59

Term	Role
143a9481	Noun
143a9481-6	Noun
146a0081	Noun
191	Noun
19r	Noun
211	Noun
231	Noun
23r	Noun
24r	Noun
25-66-01	Num
25r	Noun
261	Num
26r	Noun
28r	Noun
33-51-38	Num
500a	Noun
500b	Noun Group
500c	Noun

APPENDIX

Synonym List

Child Term	Parent Term	Term Role	Parent Role
sas	sas institute		COMPANY
51-10-02	51-40-02	Num	Num
51-40-02-0g-0	51-40-02	Num	Num
53-00-03-2r-1	51-40-02	Num	Num
51-40-05	51-40-02	Num	Num
51-40-05-0g-0	51-40-02	Num	Num
55-16-42-400-001-a	51-40-02	Num	Num
51-10-02-0g-0	51-40-02	Num	Num
51-42-06-001-001-a01	51-40-02	Num	Num
143a7505	143a7902	Noun	Noun
146a0081	143a7902	Noun	Noun
143a9481-5	143a7902	Noun	Noun
143a7505 rev	143a7902	Noun Group	Noun
bs540	bs 500d	Prop	Noun Group
bs 500a	bs 500d	Noun Group	Noun Group
bs500c	bs 500d	Noun Group	Noun Group
b737-770	bs 500d	Prop	Noun Group
bs 500b	bs 500d	Noun Group	Noun Group

Replace Table

Add Table

OK

Cancel

APPENDIX

Text parsing setting

Property	Value
Language	English
Detect	
Different Parts of Speech	Yes
Noun Groups	Yes
Multi-word Terms	SASHELP.ENGLIST
Find Entities	None
Custom Entities	
Ignore	
Ignore Parts of Speech	'Aux' 'Conj' 'Det' 'Infl'
Ignore Types of Entities	
Ignore Types of Attributes	'Num' 'Punct'
Synonyms	
Stem Terms	Yes
Synonyms	SASHELP.ENGSYMLIST
Filter	
Start List	
Stop List	SASHELP.ENGSTPLIST
Select Languages	
Report	
Number of Terms to	120000

APPENDIX - 25 TOPICS

Results - Node: 25 Topic Diagram: Incident

File Edit View Window

Topics

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.104	0.006bs,+locate,b737-700,+galley,dwg		583	10445
Multiple	2	0.089	0.006+exit,emergency,+sign,+light,light		417	9128
Multiple	3	0.078	0.006+rib,horizontal,+stab,+stabilizer,+spar		650	6094
Multiple	4	0.082	0.006+ hinge,+elevator,+arm,+stabilizer,play		310	3825
Multiple	5	0.074	0.006due,+date,due date,+tool,b737-700		718	4206
Multiple	6	0.083	0.006+engine,apu,+flight,+odor,+leak		1236	12346
Multiple	7	0.086	0.006+o-check,+card,fs,+comply		595	6573
Multiple	8	0.080	0.006+slide,+gird,+door,bar,+entry		669	8483
Multiple	9	0.098	0.006base,maintenance,+base,maintenance,aircraft,sta		524	9635
Multiple	10	0.075	0.006+type,inspection type,+inspection,sta,+floor		734	8251
Multiple	11	0.079	0.006+seat,+track,+seat track,fs,+rail		540	6052
Multiple	12	0.076	0.006+panel,+floor,cabin,53-00-50-2r-1,+verify		893	8909
Multiple	13	0.074	0.006+repair,+skin,+frame,ea,+hole		1094	11743
Multiple	14	0.076	0.006+corrective,+corrective action,+action,+ship,er		513	8121
Multiple	15	0.092	0.006+cargo,+compartment,aft cargo,fwd cargo,fwd		680	10198
Multiple	16	0.089	0.006battery,emergency,+pack,+battery pack,mop		578	9096
Multiple	17	0.095	0.006+sidewall,bs,+support,sidewall support,+limit		663	7258
Multiple	18	0.085	0.006+supply,+power,+unit,satisfactory,operational		490	6134
Multiple	19	0.079	0.006+station,+repair,xos,major,gmm		525	5106
Multiple	20	0.078	0.006+wing,+flap,+edge,+trail,outboard		739	8983
Multiple	21	0.067	0.006shear,+fitting,emx,nsa,+shear fitting		479	3607
Multiple	22	0.070	0.006assist,+bottle,+door,+door assist bottle,+service		363	4810
Multiple	23	0.076	0.006fr,+floor,+galley,shear+beam		704	9723
Multiple	24	0.077	0.006yes,+fault,+ref,+test,b737-800		1055	9364
Multiple	25	0.076	0.006lavatory,+door,sta,+floor,+grind		894	9551

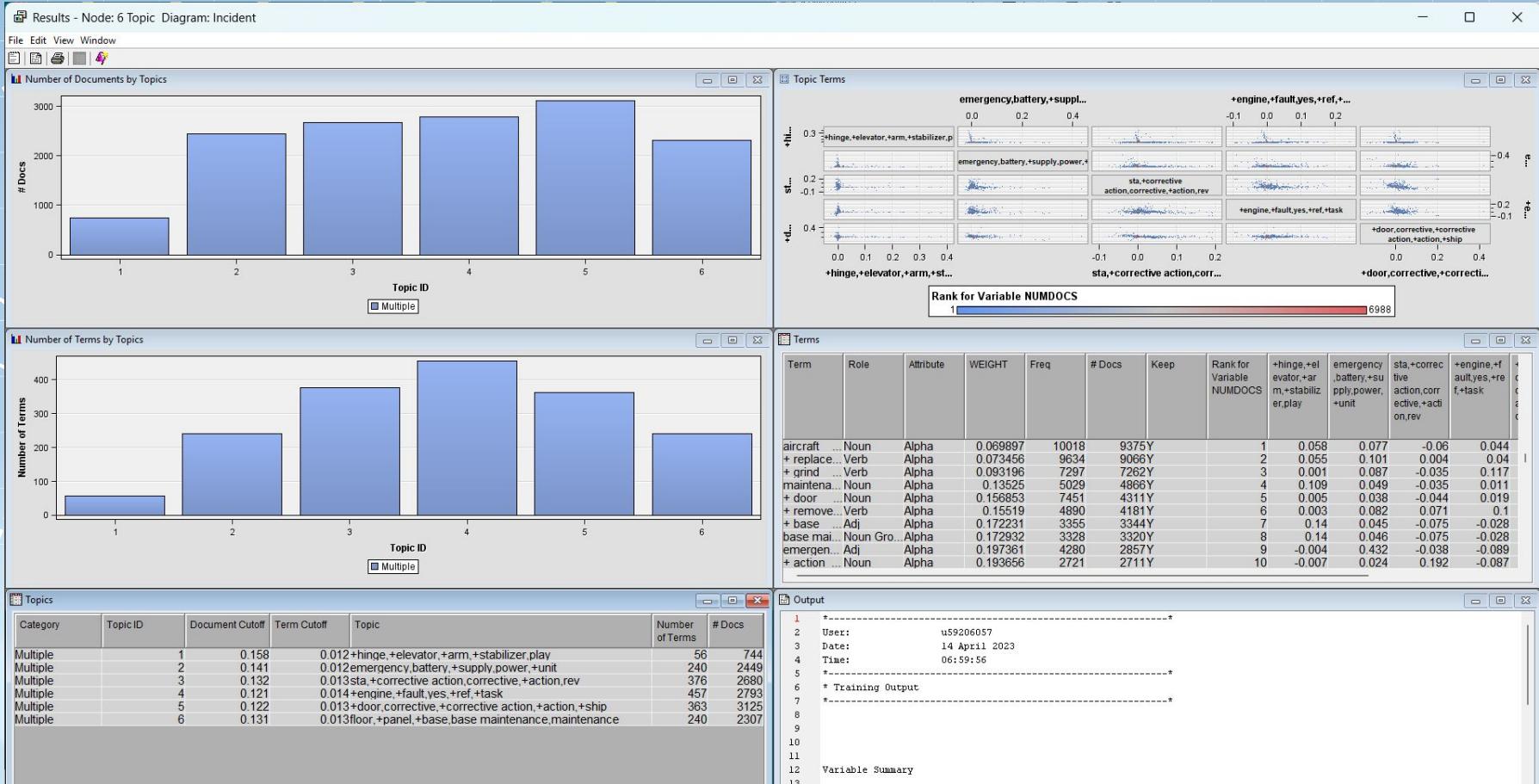
APPENDIX - 12 TOPICS

Results - Node: 12 Topic Diagram: Incident

File Edit View Window

Topics

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.122	0.012+supply,power,+unit,battery,emergency	204	1289	
Multiple	2	0.159	0.012+hinge,+elevator,+arm,+stabilizer,play	59	741	
Multiple	3	0.113	0.012+panel,floor,floor,panel,+delaminate,53-02-00-300-011	238	1605	
Multiple	4	0.110	0.013+yes,+ref,+fault,b737-800,firm	369	1904	
Multiple	5	0.116	0.013+corrective,action,corrective,+action,+ship,pri	237	2445	
Multiple	6	0.119	0.013+door,+slide,+gird,bar,+lavatory	266	2563	
Multiple	7	0.099	0.012+shear,+fitting,shear fitting,b737-57-4-1262,ssfr	169	609	
Multiple	8	0.119	0.013+seat,emergency,+exit,+light,light	261	2273	
Multiple	9	0.106	0.013+repair,+skin,rev,+perform,+crack	461	1995	
Multiple	10	0.110	0.013+sta,cargo,+corrosion,compartment,+level	244	1775	
Multiple	11	0.118	0.012+fr,+beam,srm,+galley,aft	222	1466	
Multiple	12	0.103	0.014+engine,satisfactory,+check,+flight,operational	422	2269	



APPENDIX - 6 TOPICS

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.158		0.012+hinge,+elevator,+arm,+stabilizer,play	56	744
Multiple	2	0.141		0.012emergency,battery,+supply,power,+unit	240	2449
Multiple	3	0.132		0.013sta,+corrective action,corrective,+action,rev	376	2680
Multiple	4	0.121		0.014+engine,+fault,yes,+ref,+task	457	2793
Multiple	5	0.122		0.013+door,corrective,+corrective action,+action,+ship	363	3125
Multiple	6	0.131		0.013floor,+panel,+base,base maintenance,maintenance	240	2307