

Behavioral Segmentation Model Evaluation

This week's evaluation of the behavioral segmentation models will include the two top performing models from the week 5 assignment along with a new tuned model. Each of these three models received a respectful Silhouette coefficient score around 0.7. Silhouette scores, a measure of cohesion and segmentation separation, greater than 0.5 are considered good. Scores are considered better as they approach 1. (Chorianopoulos, 2016, p. 130) From this perspective, there wasn't a clear technical segmentation winner. A deeper analysis was performed, with the detailed results included below, to select a top performing model.

The Week 5 Top Silhouette Model had a slight edge on the Silhouette score, but it was small enough that rounding the results hide this fact. This model features only 2 clusters with one cluster representing over 76% of the records. For general clustering purposes, this is typically not a desired result. A better technical results would be a handful of clusters that are not dominated with one cluster. (Chorianopoulos, 2016, p. 129) However, the clusters do have a fairly tight mean from the centroid, -0.42 and 1.15. In other words, this model may have some issues from a technical perspective, but may work well when it is evaluated with a business subject expert.

Included with all three models are some statistics on the individual clusters. These additional statistics are a starting point for evaluating the individual cluster profile. In the top silhouette model, cluster 2 is smaller but features users with a significant out call usage profile. In particular, cluster 2 customers had roughly 5 times the mobile and 3.5 times the international out call rates. A business subject matter expert may want to explore this segment for things like high value customers.

The next model, Week 5 PCA Model, produced three clusters. There was still one dominate cluster with approximately 53% of the records. Cluster 1 and 2 look solid with

segmentation means of -0.75 and 0.45. Cluster 3 showed more separation with a mean of 2.07 which can also be easily observed in the Cluster Graph. The model appears to perform better, especially if the first two clusters are of business interest. Cluster 3 may merit further analysis and segmentation depending on the business goals.

The final model, the Tuned Model, was a hand tuned K-Means model with four clusters. This model showed a number of similarities to the Week 5 PCA Model, but didn't have a single predominate cluster. Instead, this model had two clusters with roughly 35% and 33% of the records. The smallest cluster, cluster 2, was roughly the same size as the smallest Week 5 PCA Model cluster at 10%. Similarly, Cluster 2 showed a mean variation of 2.15.

A review of the Cluster Graph shows visually what has happened. Cluster 2 contains roughly the same the widely dispersed records. At the same time, a new cluster was identified and it appears to be fairly distinct. As a result, this makes the K-Means model shows greater technical strength making it the model of choice.

As mentioned, models can be evaluated from a technical perspective but the best model is the one that best matches the business goals. This will likely require working with a business subject expert and potentially augmenting the model with market research. Market research can provide additional insight into individual segments or even a way to cross-exam them.

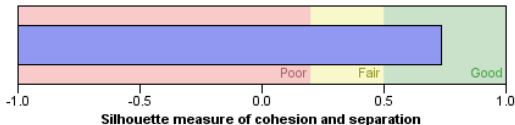
(Chorianopoulos, 2016, p. 138) In either case, picking the best model will require careful deliberation and an understanding of the technical and business trade-offs.

Week 5 Top Silhouette Model

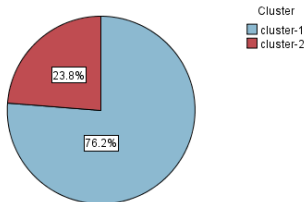
Model Summary

Algorithm	TwoStep
Inputs	1
Clusters	2

Cluster Quality



Cluster Sizes



Size of Smallest Cluster	333 (23.8%)
Size of Largest Cluster	1069 (76.2%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	3.21

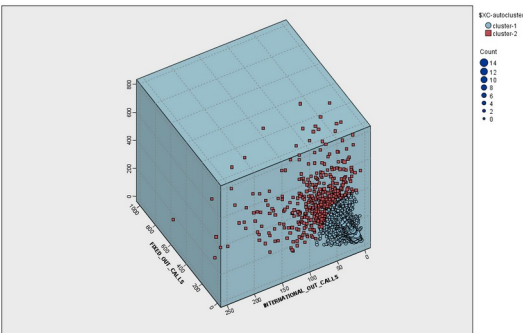
Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	cluster-1	cluster-2
Label		
Description		
Size	76.2% (1069)	23.8% (333)
Inputs	\$F-Factor-Row -0.42	\$F-Factor-Row 1.15

Cluster Graph



Cluster 1

FIXED_OUT_CALLS	
Statistics	
Count	1567
Mean	121.098
Min	0.000
Max	355.554
Standard Deviation	69.133
Median	110.260
MOBILE_OUT_CALLS	
Statistics	
Count	1567
Mean	30.582
Min	0.000
Max	301.374
Standard Deviation	41.930
Median	12.912
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	1567
Mean	11.708
Min	0.000
Max	65.662
Standard Deviation	10.376
Median	9.439

Cluster 2

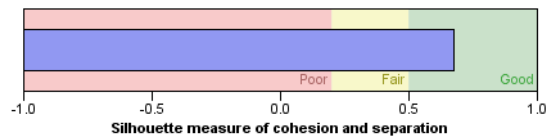
FIXED_OUT_CALLS	
Statistics	
Count	470
Mean	293.789
Min	10.116
Max	898.544
Standard Deviation	143.720
Median	281.537
MOBILE_OUT_CALLS	
Statistics	
Count	470
Mean	150.677
Min	0.151
Max	769.851
Standard Deviation	142.518
Median	116.331
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	470
Mean	40.950
Min	0.000
Max	237.254
Standard Deviation	42.948
Median	29.616

Week 5 PCA Model

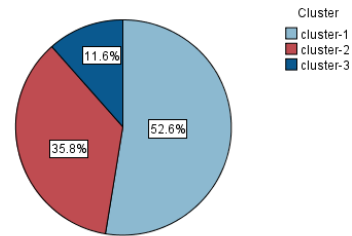
Model Summary

Algorithm	TwoStep
Inputs	1
Clusters	3

Cluster Quality



Cluster Sizes



Size of Smallest Cluster	162 (11.6%)
Size of Largest Cluster	738 (52.6%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	4.56

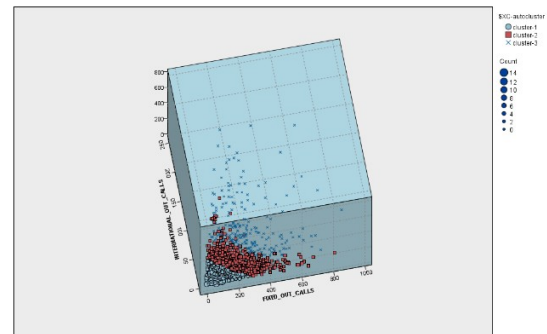
Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	cluster-1	cluster-2	cluster-3
Label			
Description			
Size	52.6% (738)	35.8% (502)	11.6% (162)
Inputs	\$F-Factor-1 -0.75	\$F-Factor-1 0.45	\$F-Factor-1 2.07

Cluster Graph



Cluster 1

FIXED_OUT_CALLS	
Statistics	
Count	1076
Mean	95.484
Min	0.000
Max	264.755
Standard Deviation	50.513
Median	90.286
MOBILE_OUT_CALLS	
Statistics	
Count	1076
Mean	17.770
Min	0.000
Max	165.794
Standard Deviation	23.519
Median	7.955
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	1076
Mean	8.655
Min	0.000
Max	37.904
Standard Deviation	6.736
Median	7.805

Cluster 2

FIXED_OUT_CALLS	
Statistics	
Count	734
Mean	211.034
Min	10.116
Max	824.324
Standard Deviation	108.977
Median	193.309
MOBILE_OUT_CALLS	
Statistics	
Count	734
Mean	79.742
Min	0.000
Max	752.718
Standard Deviation	97.646
Median	49.950
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	734
Mean	19.151
Min	0.000
Max	139.148
Standard Deviation	17.460
Median	15.538

Cluster 3

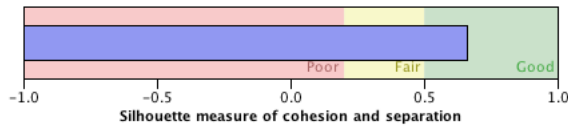
FIXED_OUT_CALLS	
Statistics	
Count	227
Mean	309.260
Min	63.188
Max	898.544
Standard Deviation	149.639
Median	288.179
MOBILE_OUT_CALLS	
Statistics	
Count	227
Mean	181.012
Min	0.817
Max	769.851
Standard Deviation	139.327
Median	158.879
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	227
Mean	62.659
Min	0.000
Max	237.254
Standard Deviation	48.122
Median	51.448

Tuned Model

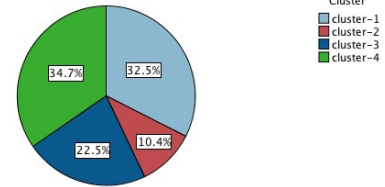
Model Summary

Algorithm	K-Means
Inputs	1
Clusters	4

Cluster Quality



Cluster Sizes



Size of Smallest Cluster	146 (10.4%)
Size of Largest Cluster	486 (34.7%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	3.33

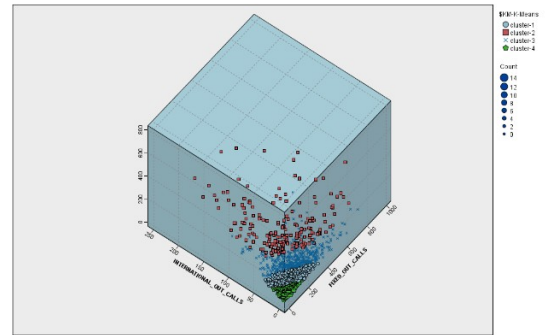
Clusters

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	cluster-4	cluster-1	cluster-3	cluster-2
Label				
Description				
Size	34.7% (486)	32.5% (455)	22.5% (315)	10.4% (146)
Inputs	DT-Factor-1 -0.94	DT-Factor-1 -0.20	DT-Factor-1 0.77	DT-Factor-1 2.13

Cluster Graph



Cluster 1

FIXED_OUT_CALLS	
Statistics	
Count	671
Mean	154.741
Min	10.116
Max	352.777
Standard Deviation	62.067
Median	150.175
MOBILE_OUT_CALLS	
Statistics	
Count	671
Mean	34.631
Min	0.000
Max	495.141
Standard Deviation	39.568
Median	21.073
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	671
Mean	13.501
Min	0.000
Max	46.824
Standard Deviation	9.136
Median	12.042

Cluster 2

FIXED_OUT_CALLS	
Statistics	
Count	206
Mean	310.779
Min	63.188
Max	898.544
Standard Deviation	149.738
Median	288.978
MOBILE_OUT_CALLS	
Statistics	
Count	206
Mean	189.802
Min	0.817
Max	769.851
Standard Deviation	141.456
Median	180.346
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	206
Mean	64.851
Min	0.000
Max	237.254
Standard Deviation	49.229
Median	53.390

Cluster 3

FIXED_OUT_CALLS	
Statistics	
Count	457
Mean	236.100
Min	28.542
Max	824.324
Standard Deviation	126.159
Median	207.467
MOBILE_OUT_CALLS	
Statistics	
Count	457
Mean	104.060
Min	0.151
Max	752.718
Standard Deviation	112.521
Median	72.202
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	457
Mean	22.698
Min	0.000
Max	139.148
Standard Deviation	21.324
Median	18.636

Cluster 4

FIXED_OUT_CALLS	
Statistics	
Count	703
Mean	74.101
Min	0.000
Max	188.424
Standard Deviation	37.463
Median	70.615
MOBILE_OUT_CALLS	
Statistics	
Count	703
Mean	12.587
Min	0.000
Max	122.913
Standard Deviation	17.105
Median	5.516
INTERNATIONAL_OUT_CALLS	
Statistics	
Count	703
Mean	6.831
Min	0.000
Max	26.275
Standard Deviation	5.287
Median	6.823

References

Chorianopoulos, A. (2016). *Effective CRM Using Predictive Analytics*. West Sussex, United Kingdom: John Wiley & Sons, Ltd.