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

|  |  |
| --- | --- |
|  |  |
|  | Cluster Graph |
| Cluster 1 | Cluster 2 |

Week 5 PCA Model

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | |
|  | | Cluster Graph | |
| Cluster 1 | Cluster 2 | | Cluster 3 |

Tuned Model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | |  | |
|  | | | Cluster Graph | |
| Cluster 1 | Cluster 2 | Cluster 3 | | Cluster 4 |

# References

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