Samsung Galaxy S7 Edge Review Analysis

Jonah Turner

INFO 4670

9th December, 2020

Review Data

I collected data on reviews of the Samsung Galaxy S7 Edge from a Kaggle dataset called 1.4 million cell phone reviews collected by the user Zootojia. This dataset contained reviews on many different phones, so the Samsung Galaxy S7 was selected to reduce the size of the dataset to a manageable level. The data was further cleaned to remove missing values in the operations.

Analysis Goal

The goal of the analysis is to find the features of the phone that are most strongly related to positive ratings. These could include the camera, the edge design, the battery, and several other common features of modern smartphones.

Association Analysis

The results of the association analysis reveal some of the most important features of the phone in determining satisfaction. The association graph, including only results with relatively high confidence, reveals that “camera” and “battery life” are influential in customer experience with the phone. The association data includes these features but adds importance to the “edge” and “screen”. In general, the association data portrays a positive association in the reviews, with no results in the most supported results containing words with a negative sentiment.

A picture containing chart

Description automatically generated

Diagram

Description automatically generated

Clustering Analysis

A k-means clustering analysis shows words that were commonly associated in different reviews. In deciding the number of clusters to include, three different numbers of clusters were tested: 10, 6, and 4. Reducing the number of clusters increased the definition of the clusters and removed several clusters that performed weakly, so the analysis utilizes the algorithm with four centroids.

Average within-centroid distance and the Davies-Bouldin index suffered slightly with the decreasing of centroids, but as the analysis improved overall with the decreasing of the number of centroids, the loss of performance in these two metrics was acceptable. The final four clusters had average within-centroid distances ranging between .8 and .99 and produced a Davies-Bouldin index of 7.257.

The clusters reveal patterns among reviews. Cluster 0 shows a strong relationship with the word “product” and weakly included several words like “excellent”, “great”, “Samsung”, and “recommend”. As in the association analysis, no negative sentiments were included in the strongest results of cluster 0.

Cluster 1 shows a strong relationship with the words “love” and “great”, showing a positive sentiment overall for the cluster, and includes words with weak strengths such as “features”, “phone”, “camera”, and “edge”. This cluster connects some of the features with positive sentiments and begins to highlight the features most important to customer sentiment.

Cluster 2 shows relatively weak strengths across all words, with the strongest words including “phone”, “edge”, “Samsung”, “galaxy”, “great”, and “screen”. This cluster includes features not found in cluster 1, which may indicate that these features are less important to customer satisfaction than those in cluster 1.

Cluster 3 shows an association of moderate strength with the words “battery” and “life”. The inclusion of these words as the strongest words of the cluster, followed immediately by “great” with a relatively weak strength, suggest the importance of battery life among consumers. Other words ranked highly in cluster 3 include “camera”, “good”, “fast”, and “phone”. This cluster gives further insight into the features most highly valued by customers.

Performance Vector:

Text, letter

Description automatically generated

Cluster data (Cluster 3 as example):

A picture containing table

Description automatically generated

Conclusion

Both the association analysis and the clustering analysis highlighted specific features that were associated with positive customer sentiment. Commonly praised features of the Samsung Galaxy S7 Edge include its battery life, camera, edge, and screen. It is important to remember when conducting the analysis that while the phone’s name includes “Edge”, the edge of the phone is also an important feature. Even with the inclusion of “edge” in the name of the phone, however, “edge” was not as strongly associated with user sentiment as other features of the phone. This may be a useful consideration for Samsung’s naming convention, as the feature they may have assumed would be most important to consumers as to include it in the name of the product was not as important as more common smartphone tropes.

An interesting phenomenon in both analyses was the complete lack of negative sentiments. The data was not cleaned in a way to favor positive reviews. It more strongly suggests a bias in customers that choose to leave a review. These analyses are also designed so as to not show the relatively low number of negative reviews. An analysis of negative reviews would be better conducted using different methods or including only negative reviews in the dataset.