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.

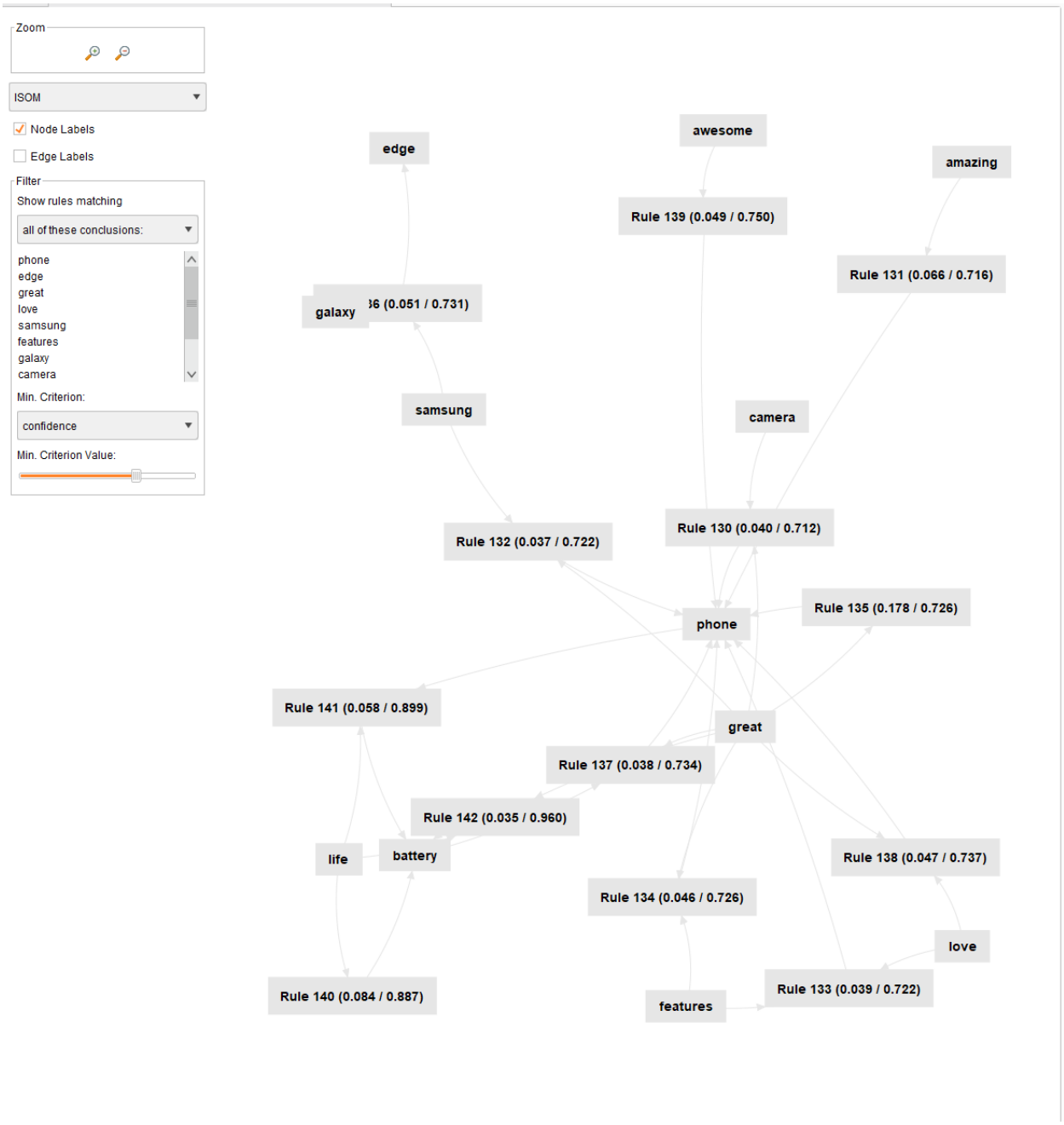
Analysis Goal

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.

No.	Premises	Conclusion	Support ↓	Confidence	La
45	phone	great	0.178	0.280	0.7
135	great	phone	0.178	0.726	0.9
43	phone	edge	0.175	0.275	0.7
101	edge	phone	0.175	0.594	0.9
35	phone	love	0.166	0.262	0.7
126	love	phone	0.166	0.692	0.9
107	samsung	phone	0.142	0.629	0.9
123	features	phone	0.112	0.686	0.9
70	edge	galaxy	0.104	0.352	0.8
120	galaxy	edge	0.104	0.683	0.9
68	edge	samsung	0.103	0.349	0.8
89	samsung	edge	0.103	0.456	0.9
124	camera	phone	0.098	0.687	0.9
118	battery	phone	0.096	0.681	0.9
119	screen	phone	0.096	0.682	0.9
98	galaxy	phone	0.088	0.583	0.9
57	edge	love	0.088	0.300	0.8
72	love	edge	0.088	0.367	0.8
100	battery	life	0.084	0.592	0.9
140	life	battery	0.084	0.887	0.9
27	edge	great	0.073	0.248	0.8
56	great	edge	0.073	0.297	0.8
59	samsung	galaxy	0.070	0.312	0.8
91	galaxy	samsung	0.070	0.464	0.9
129	fast	phone	0.067	0.704	0.9
131	amazing	phone	0.066	0.716	0.9
81	features	edge	0.065	0.402	0.9
117	life	phone	0.064	0.679	0.9
34	great	love	0.064	0.262	0.8
41	love	great	0.064	0.267	0.8
37	love	samsung	0.064	0.265	0.8
50	samsung	love	0.064	0.283	0.8
111	good	phone	0.063	0.647	0.9
33	great	features	0.063	0.257	0.8
77	features	great	0.063	0.387	0.9
88	screen	edge	0.062	0.436	0.9


Zoom
ISOM
Node Labels
Edge Labels
Filter
Show rules matching
all of these conclusions:
phone
edge
great
love
samsung
features
galaxy
camera
Min. Criterion:
confidence
Min. Criterion Value:
edge
awesome
amazing
Rule 139 (0.049 / 0.750)
Rule 131 (0.066 / 0.716)
galaxy
samsung
camera
Rule 132 (0.037 / 0.722)
Rule 130 (0.040 / 0.712)
Rule 135 (0.178 / 0.726)
phone
Rule 141 (0.058 / 0.899)
great
Rule 137 (0.038 / 0.734)
Rule 142 (0.035 / 0.960)
life
battery
Rule 138 (0.047 / 0.737)
Rule 134 (0.046 / 0.726)
love
Rule 140 (0.084 / 0.887)
features
Rule 133 (0.039 / 0.722)

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:

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -0.962
Avg. within centroid distance_cluster_0: -0.845
Avg. within centroid distance_cluster_1: -0.902
Avg. within centroid distance_cluster_2: -0.983
Avg. within centroid distance_cluster_3: -0.921
Davies Bouldin: -7.257
```

Cluster data (Cluster 3 as example):

Attribute	cluster_0	cluster_1	cluster_2	cluster_3 ↓
battery	0.004	0.004	0.005	0.163
life	0.002	0.002	0.002	0.144
great	0.049	0.119	0.023	0.064
camera	0.017	0.051	0.015	0.051
good	0.042	0.012	0.021	0.044
fast	0.010	0.031	0.014	0.038
phone	0.016	0.063	0.035	0.034
love	0.030	0.211	0.016	0.033
edge	0.016	0.051	0.029	0.031
amazing	0.018	0.029	0.017	0.029
performance	0.005	0.007	0.009	0.029
features	0.029	0.091	0.016	0.027
screen	0.005	0.021	0.022	0.026
excellent	0.056	0.011	0.009	0.026
takes	0.001	0.021	0.004	0.025
lasts	0.003	0.001	0.001	0.024
light	0.003	0.008	0.005	0.024
display	0.007	0.012	0.009	0.022
awesome	0.019	0.046	0.011	0.022
pictures	0.004	0.025	0.006	0.021
galaxy	0.013	0.039	0.024	0.021
water	0.004	0.015	0.008	0.021
seen	0.001	0.002	0.003	0.020
samsung	0.046	0.048	0.029	0.019
match	0	0	0.001	0.018
rear	0	0.001	0.000	0.018
charge	0.001	0.004	0.004	0.018
longer	0.001	0.002	0.002	0.018
design	0.015	0.037	0.012	0.017
stunning	0	0.002	0.002	0.017
nice	0.015	0.020	0.013	0.016
able	0	0.003	0.004	0.016
easy	0.022	0.047	0.012	0.015
charging	0.001	0.005	0.003	0.015

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.