Which Smart Phone to Choose?

BY: MOHANA KRISHNA CHAITANYA KORLEPARA
GRADUATE STUDENT
MASTER'S IN BUSINESS ANALYTICS

Abstract

Android smartphones have over 80% of the market share in the smartphone industry. With so many new phones launched with similar price points and features it almost makes it impossible for a customer to make a decision and be satisfied. One effective approach to this problem is by using the experiences of users of these products to draw insights and reach a conclusion. Experiences of these customers are best captured in the form of feedback or reviews. What can be a better source of such reviews than the largest online marketplace, Amazon. Reviews are individual perspectives, which are very diverse and cover both positive and negative emotions of customers for a product. Analyzing the details of these reviews could provide more information than just plain specifications of smartphones. Information regarding performance of the touchscreen, actual abilities of the camera, music and audio experiences, and other important factors could be insightful for the phone buyer as well as phone manufacturer. Buyers can narrow down their choice to a single product based on their most desired aspect of a phone. Whereas, manufacturers can understand the limitations of their current version and come up with a better product to have success in this competitive and growing marketspace.

In this paper, my objective was to analyze the overall sentiment of the reviews of smartphones, which fall in similar price points. In order to pursue this, I have extracted the user comments from Amazon using a web crawler and used SAS Enterprise Miner and SAS Sentiment Studio to draw insights.

Introduction

One of the largest online market places in the world is Amazon; it has a huge and diverse customer base all over the world. Reviews of the customers on the unlocked smartphones from Amazon should be a great place to start analyzing the smartphone customer's reviews.

Using text mining, the most frequently used terms in the user reviews could be extracted giving the sense of like and dislikes of the user on a given phone. We can further analyze the terms and find how well they are associated with other words. Analyzing these parameters would help us measure the customer satisfaction and dissatisfaction and how the product is doing in the market. Using this analysis, phone manufacturers can understand their competitors and find the improvements that could be implemented in the next release to improve customer satisfaction as well as sales of the product.

Dataset

The data set for this analysis contains the Product Title, Brand, Price, Rating, Reviews of smartphones, which were collected from the Amazon website. The data was taken from the https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones. The dataset had over 400,000 customer reviews on various phones.

Data Dictionary

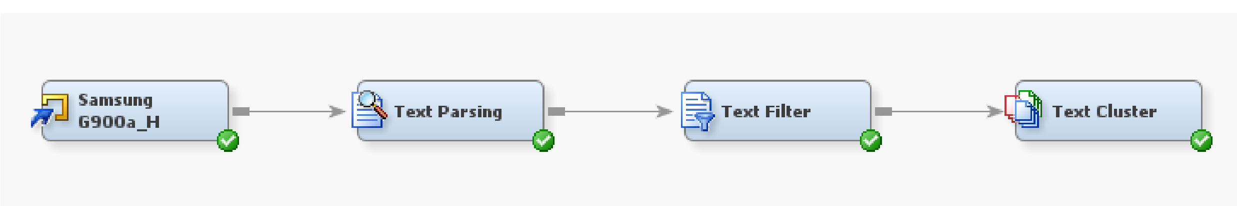
Variable	Level	Descriptor
Product Title	Nominal	Name of the product
Brand	Nominal	Name of the brand
Price	Interval	Price of the phone
Ratings	Interval	The rating of the phone
Review	Text	The review of the phone given by the customer

Table1: Table dictionary

Approach

The dataset consists of the over 400,000 reviews on smartphones from different customers all over the world. For the purpose of this analysis, a subset of the data was taken consisting of two phone brands, Samsung and BLU, which have products that were in the same price range of $179-229. The phones considered for this project were Samsung Galaxy S5 G900A and BLU Vivo5.

Methodology



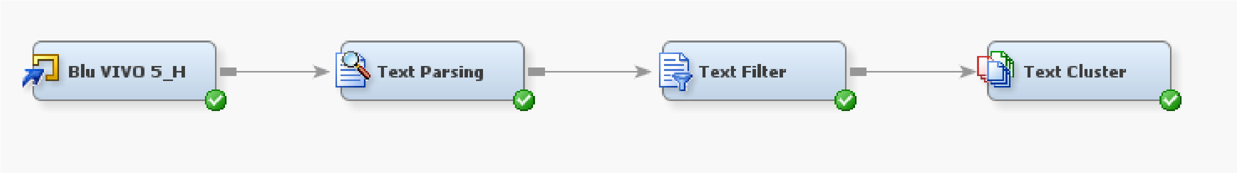


Fig 1: Text mining Process

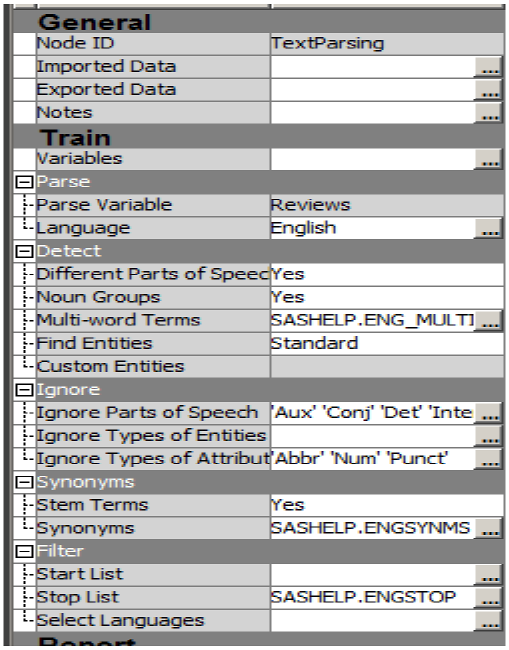
Text Import

The Text Import node was used to import the dataset. Below are the settings that were used for importing the data.

Property	Value
General	
Node ID	FIMPORT
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Import File	M:\mkorlep\paper\sa ...
Maximum rows to import	1000000
Maximum columns to imp	10000
Delimiter	,
Name Row	Yes
Number of rows to skip	0
Guessing Rows	500
File Location	Local
File Type	xlsx
Advanced Advisor	No
Rerun	No

Text Parsing

The Text Parsing node was used to clean the data by removing redundant and irrelevant data. The following properties were used for the Text Parsing node:



From the text parsing, we also obtain the frequency matrix that shows the number of times certain variables repeat themselves. This information gives an understanding of which variables are important.

Term	Role	Attribute	Freq	# Docs	Keep	Parent/Child Status	Parent ID	Rank for Variable numdocs
+ phone ...	Noun	Alpha	510	282	Y	+	727	1
+ be ...	Verb	Alpha	468	212	N	+	2800	2
+ have ...	Verb	Alpha	263	138	N	+	2746	3
not ...	Adv	Alpha	183	114	N		2755	4
+ work ...	Verb	Alpha	126	108	Y	+	1810	5
great ...	Adj	Alpha	116	103	Y		359	6
very ...	Adv	Alpha	104	94	N		2731	7
+ new ...	Adj	Alpha	123	92	N	+	2826	8
+ love ...	Verb	Alpha	86	84	Y	+	275	9
+ do ...	Verb	Alpha	122	80	N	+	2835	10
+ good ...	Adj	Alpha	83	78	Y	+	154	11
+ get ...	Verb	Alpha	99	68	N	+	2712	12
excellent ...	Adj	Alpha	59	58	Y		1054	13
no ...	Adv	Alpha	72	58	N		2850	13
+ use ...	Verb	Alpha	66	58	N	+	2750	13
great ...	Noun	Alpha	56	54	Y		1942	16
so ...	Adv	Alpha	70	50	N		2770	17
+ buy ...	Verb	Alpha	57	47	Y	+	81	18
great phon...	Noun Group	Alpha	49	47	Y		1026	18
+ product ...	Noun	Alpha	45	45	Y	+	694	20
+ come ...	Verb	Alpha	50	44	N	+	2856	21
s ...	Noun	Alpha	50	42	N		2710	22
happy ...	Adj	Alpha	43	41	Y		1560	23
+ price ...	Noun	Alpha	44	41	Y	+	2049	23
+ problem ...	Noun	Alpha	46	40	Y	+	682	25
samsung ...	Prop	Alpha	46	39	Y		2089	26
far ...	Adv	Alpha	44	37	Y		1607	27
+ seller ...	Noun	Alpha	50	37	Y	+	232	27
+ unlock ...	Verb	Alpha	52	36	Y	+	894	29
s5 ...	Prop	Mixed	41	34	Y		1171	30
t ...	Noun	Alpha	46	34	N		2771	30

Fig 2: Text parsing output

Text Filter

By using the text filter, we try to eliminate the terms that are least frequent in all the reviews.

Term	Role	Attribute	Status	Weight	Imported Frequency	Freq	Number of Imported Documents	# Docs	Rank	Parent/Child Status	Parent ID
+ phone ...	Noun	Alpha	Attribute	0.157	501	502	273	274	1	+	799
+ be ...	Verb	Alpha	Drop	0.000	468	468	212	212	2	+	2984
+ have ...	Verb	Alpha	Drop	0.000	263	263	138	138	3	+	2930
+ not ...	Adv	Alpha	Drop	0.000	183	184	114	114	4	+	2939
+ work ...	Verb	Alpha	Keep	0.273	126	126	108	108	5	+	1968
+ great ...	Adj	Alpha	Keep	0.270	116	120	103	107	6	+	406
+ very ...	Adv	Alpha	Drop	0.000	104	105	94	95	7	+	2915
+ new ...	Adj	Alpha	Drop	0.000	123	123	92	92	8	+	3010
+ love ...	Verb	Alpha	Keep	0.303	86	87	84	84	9	+	311
+ good ...	Adj	Alpha	Keep	0.309	83	87	78	82	10	+	178
+ excelle...	Adj	Alpha	Keep	0.309	59	83	58	81	11	+	1140
+ do ...	Verb	Alpha	Drop	0.000	122	122	80	80	12	+	3019
+ get ...	Verb	Alpha	Drop	0.000	99	99	68	68	13	+	2896
+ use ...	Verb	Alpha	Drop	0.000	66	68	58	60	14	+	2934
+ no ...	Adv	Alpha	Drop	0.000	72	73	58	59	15	+	3035
great ...	Noun	Alpha	Keep	0.373	56	56	54	54	16		2119
so ...	Adv	Alpha	Drop	0.000	70	70	50	50	17		2954
+ buy ...	Verb	Alpha	Keep	0.408	57	57	47	47	18	+	98
great ph...	Noun Gr...	Alpha	Keep	0.397	49	49	47	47	18		1112
+ sams...	Company	Entity	Keep	0.409	46	53	39	45	20	+	570
+ produc...	Noun	Alpha	Keep	0.400	45	45	45	45	20	+	765
+ come ...	Verb	Alpha	Drop	0.000	50	51	44	45	20	+	3041
+ s ...	Noun	Alpha	Drop	0.000	50	51	42	43	23	+	2895
happy ...	Adj	Alpha	Keep	0.417	43	43	41	41	24		1686
+ price ...	Noun	Alpha	Keep	0.418	44	44	41	41	24	+	2230
+ proble...	Noun	Alpha	Keep	0.430	46	47	40	40	26	+	751
+ far ...	Adv	Alpha	Keep	0.437	44	45	37	38	27	+	1745
+ seller ...	Noun	Alpha	Keep	0.447	50	50	37	37	28	+	266
+ unlock...	Verb	Alpha	Keep	0.448	52	52	36	36	29	+	979
t ...	Noun	Alpha	Drop	0.000	46	46	34	34	30		2955
+ card ...	Noun	Alpha	Keep	0.469	33	39	28	32	31	+	555
+ thing ...	Noun	Alpha	Keep	0.455	33	33	32	32	31	+	1242
+ brand ...	Noun	Alpha	Keep	0.454	32	32	32	32	31	+	1618

Fig 3: Text Filter Output

From the text filter, we can drop the words that are not important for the analysis, as they do not imply performance of smartphone in any way. Only the words that describe the phone are kept. Also, synonymous words are grouped together to avoid the redundancy in the analysis.

Concept Links

For the Samsung Galaxy G900a:

Concept links were found in the interactive filter of the Text Filter node in SAS Enterprise Miner. It shows the association between the words used to give the review. The width of the connecting line in a concept link indicates the strength of association.

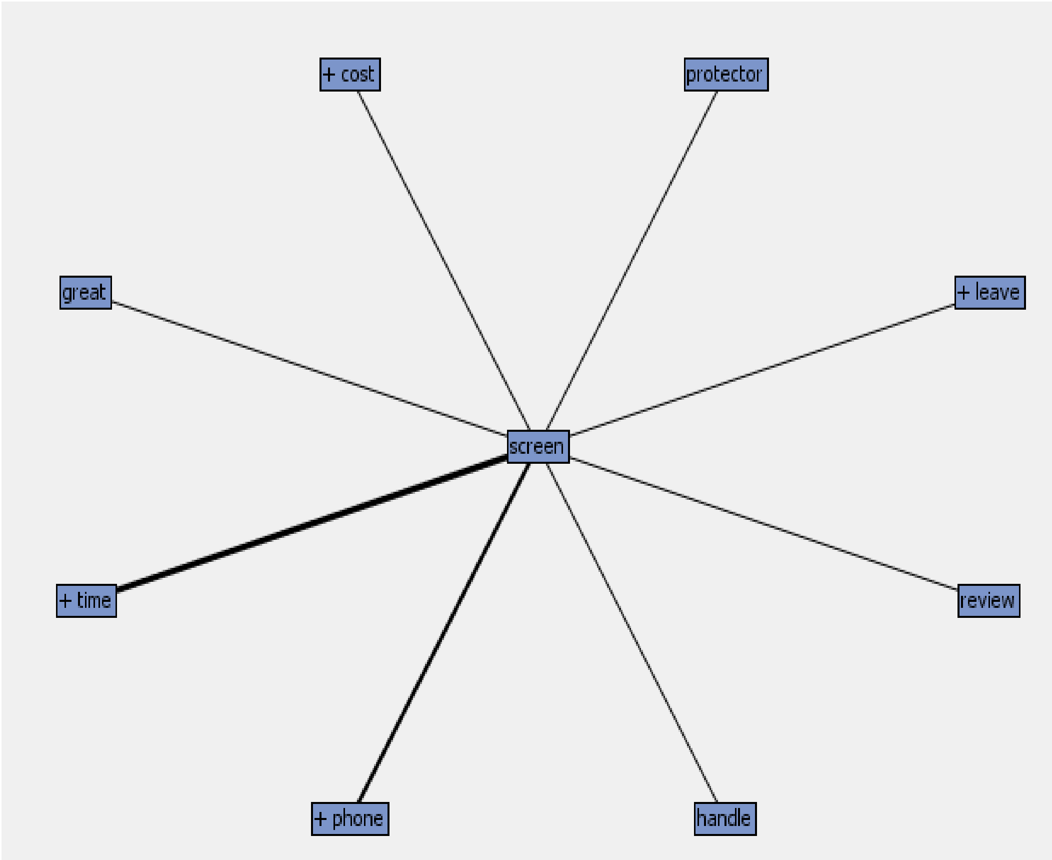


Fig 4: Concept link for a word in low rated Samsung Galaxy G900.

From the reviews, it can be seen that the screen has highest association with the time, we can interpret this as the screen response time being longer causing discomfort among the user and resulting in a poor review.

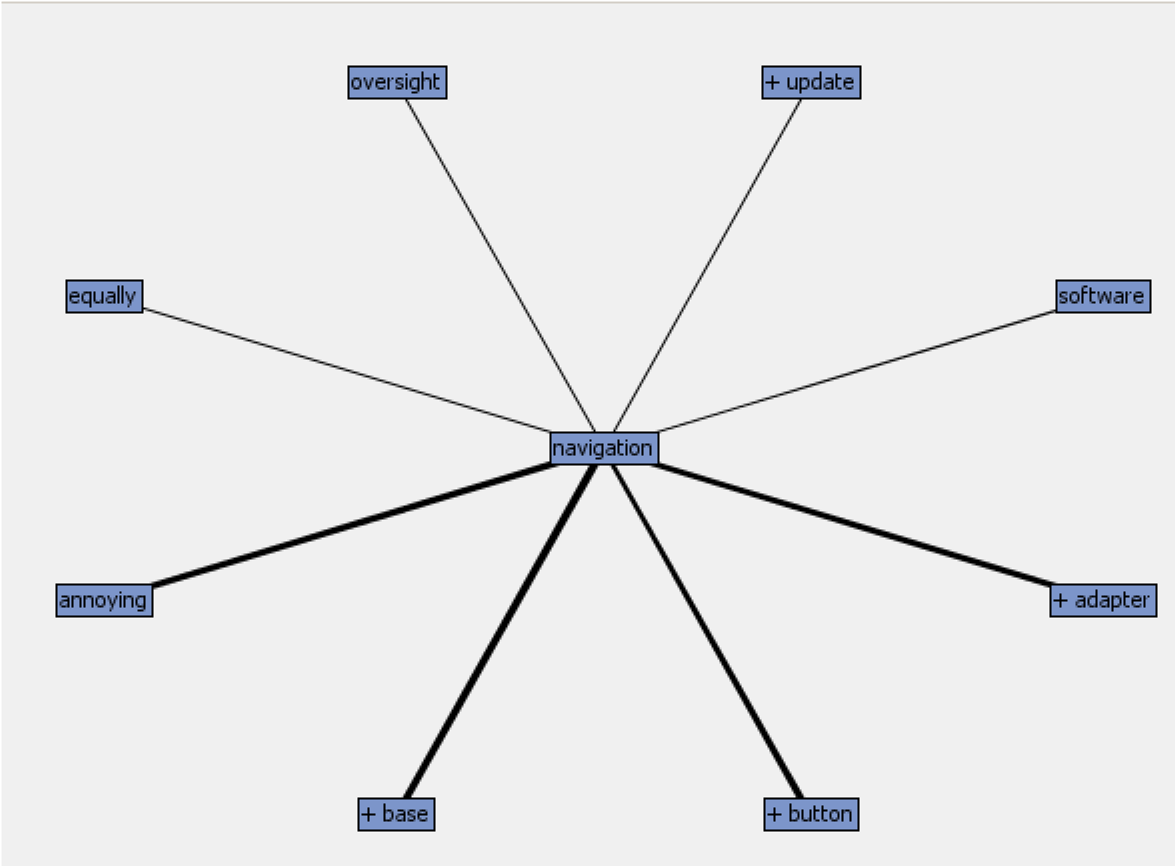


Fig 5: Concept link negative reviews of BLU VIVO5

Navigation has a high association with annoying suggesting that the phone's inbuilt software is causing problems to the users in navigation hence making them give negative reviews/comments about the phone.

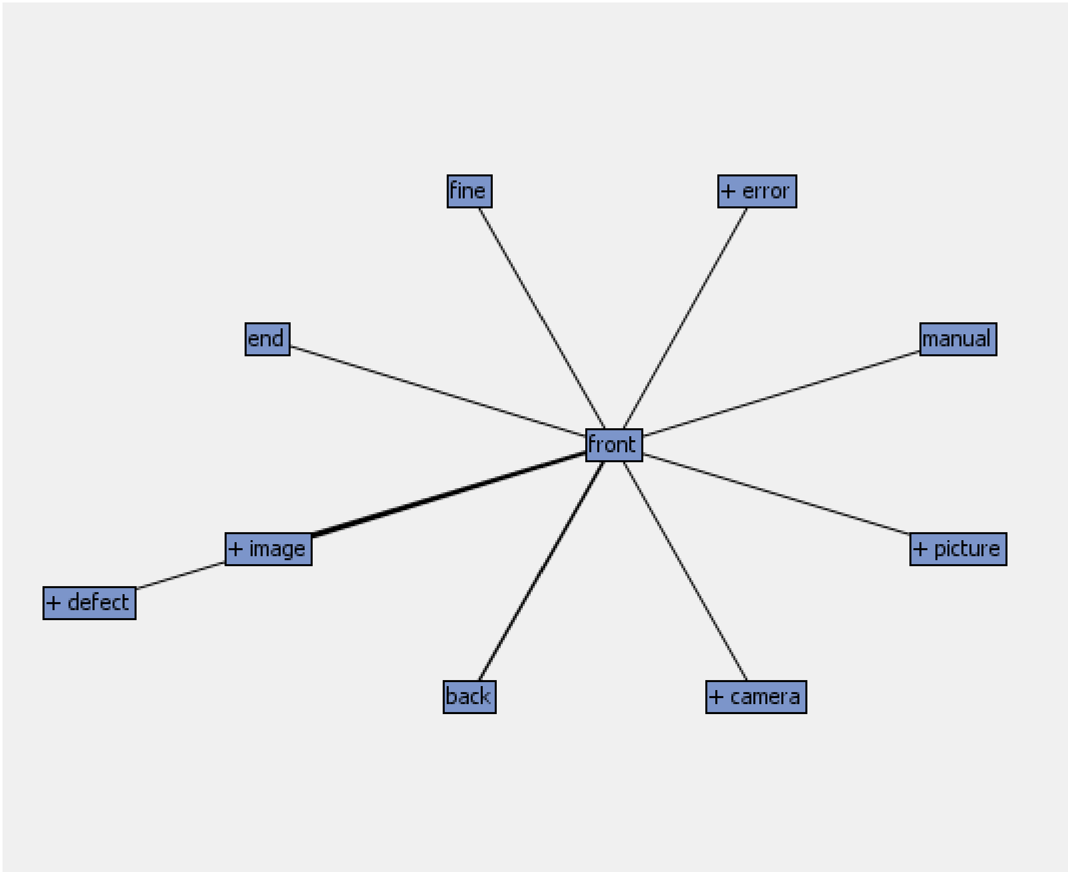


Fig 6 : Concept link of the negative reviews in BLU VIVO5

From the concept link in figure 6, we can observe that many customers have faced problems in the front camera and on further analysis down the tree we can see that there are defects which is causing the customers to write a negative comment/ review for the phone.

Text Clustering

The text clustering can be performed by text filter node in Enterprise Miner which groups together the similar terms or terms that are used together and fall into one category

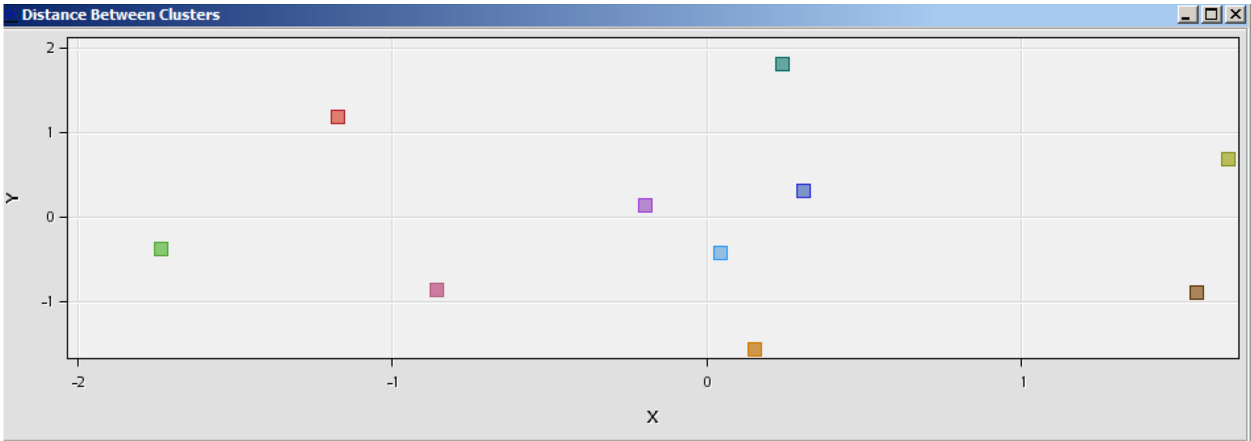


Fig 7: Distance between the clusters of the Samsung phone review

There exactly 10 clusters which are generated by the Text Clustering node, the descriptive terms that have low distances are grouped together.

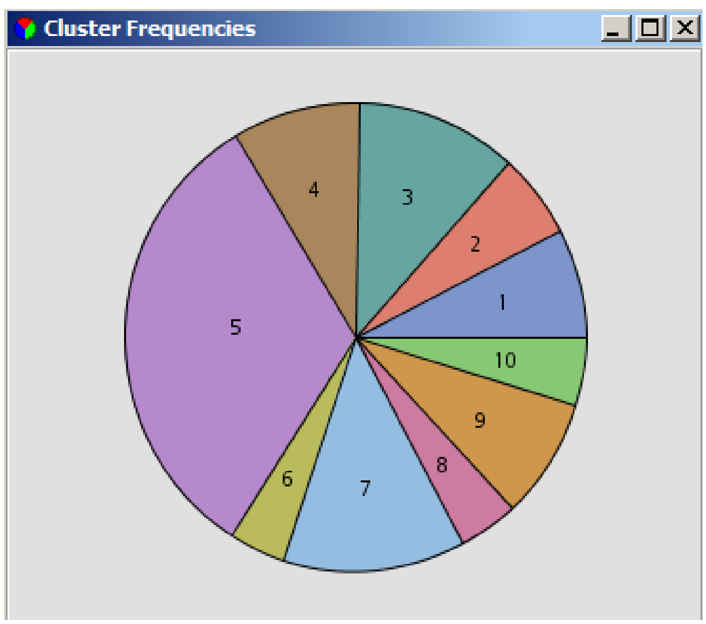


Fig 8: Cluster Distributions for the Samsung galaxy G900 reviews

Clusters

Cluster ID	Descriptive Terms	Frequency	Percentage
1	excelente ...	43	7%
2	+good +'good condition' +condition 'great condition' 'good phone' gs5 'as well' little +button +run 'go...	35	6%
3	excellent +product excelent 'excellent product' 'excellent quality' 'great product' delivery +satisfy custo...	66	11%
4	great 'great phone' +price 'great price' +phone 'good price' +carry +picture +generation +impress ab...	52	9%
5	samsung +brand +card +purchase sim +carrier +receive +unlock s5 +charge first +package +deli...	189	32%
6	+expectation +meet perfectly +side item +function yes +work +damage +sign wonderful clear +res...	23	4%
7	apps +nice android +upgrade +easy +battery +problem +issue +big +thing case +update +setting ...	74	13%
8	'great buy' +'well phone' buy dad far best +buy quality wonderful good few clear able +week +ship ...	25	4%
9	+work fine husband super great well +look +amaze +scratch +tower +love +lot straight talk +allo...	48	8%
10	+love christmas +text read 'awesome phone' life battery camera +long +year +old +receive +keep ...	28	5%

Fig 9: Cluster description for the Samsung galaxy G900 reviews

From the above cluster description, we find that there is cluster with only one word with the percentage of 7% which shows this group liked the phone a lot. Similarly, we get a cluster accounting for 32% of the records, which is highly influenced by the Samsung brand, carrier service, package and other details. A cluster with 11% feels that the product is of very high quality.

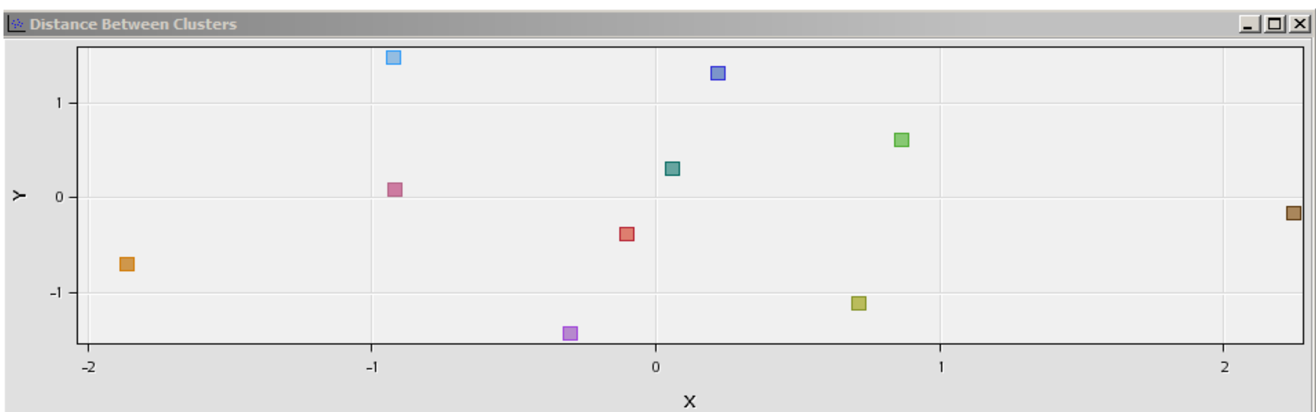


Fig 10: Cluster distances for the BLU VIV0 5 reviews

The clusters are well distributed with good distance spacing between them.

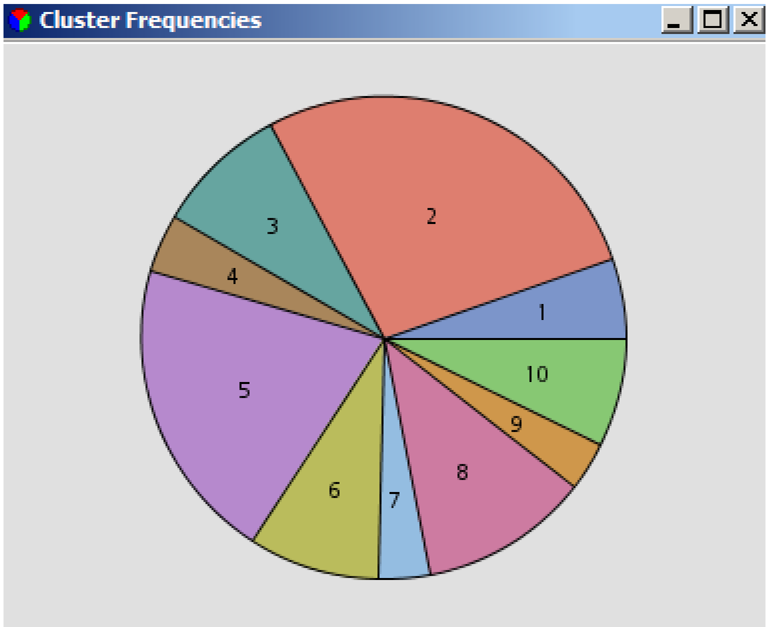


Fig 11: Cluster description for the BLU VIV0 5 reviews

Clusters

Cluster ID	Descriptive Terms	Frequency	Percentage
1	excelente +gb +notice nexus premium +ram amoled os +apple +android +fingerprint +browser +device +app. +spe...	76	5%
2	good galaxy samsung +screen +feature +expectation +recommend pretty +nice +flagship +amaze +cost +camera +t...	398	27%
3	gold +color fine +look +freeze +happy back beautiful super +speaker +week +android amazon apps +work ...	131	9%
4	excellent 'excellent product' +product 'excellent build quality' 'excellent phone' +compete +compensate +freeze graphics...	55	4%
5	+phone +price +good +great +love +'great phone' awesome +'well phone' 'awesome phone' money best quality buy ...	294	20%
6	vivo blu +product xl +smartphone cell +spend +brand +cellphone +want +cost +device +great +price +flagship ...	131	9%
7	value money size +game processor performance little +allow +device +low well +light +video +notice +amaze ...	47	3%
8	+card sim sd +problem +work far +update os +day battery +issue +app. feel +time fast ...	165	11%
9	excelent +expect great +depend moderate heavy feel +band +flagship home galaxy overall premium better 'battery li...	47	3%
10	+network +nice 'nice phone' love sim first right +charge +card +allow buy +thing galaxy +case worth ...	106	7%

Fig 9: Cluster description for the BLU vivo 5

According to the cluster description for the BLU Vivo 5, we find that there is cluster2 with 27% contribution to the total which compares the phone to Samsung galaxy, they like the screen, cost of the product. Cluster 5, with 20% contribution, feels that it's an awesome phone at great price. Cluster 7 feels that the processor performance is a bit low for the playing games

Results

1) We could find few recurring issues that were causing customer dissatisfaction leading to a poor review.
2) In the Samsung galaxy G900, the screen is strongly associated with time suggesting the response time of the screen is low
3) Navigation on the BLU VIVO 5 was troublesome and caused discomfort among the users.
4) In the BLU VIVO 5, there are many common issues in front Camera.
5) From the clusters, we could find the good attributes that were mentioned in the reviews.

Future Scope

This analysis was limited to only single of models of Samsung and Blu that were in the same price point but it gave insights on the few common issues and troubles that were observed. This analysis can be further applied to different models from different segments to determine the best value. Similarly, this analysis can be applied to other electronic gadgets like laptops so as to increase customer satisfaction and also improve the brand value of the company.

Conclusions

A major source of information for the company's to get the feedback on their product is through reviews. The reviews were in the form of text, which is an unstructured data, and more than 75% of data that was available is in the form of unstructured data. The Phone reviews were an important feedback channel for company's to improve their product and cope with the stiff competition they face. This information can be leveraged to make improvements in the upcoming models and learn what people feel about their competitors. They quickly learn from the common occurring problem to have a good reputation for the brand value they carry and also retain their customers, like in BLU Vivo 5 there were many complaints on the front camera, if these issues are detected early it can be rectified. Similarly, this analysis is useful for people who are looking to buy new phones it can help them compare the pros and cons of models beyond just the technical specifications of phones.

References

1. http://www.kdnuggets.com/2017/01/data-mining-amazon-mobile-phone-reviews-interesting-insights.html
2. https://www.educba.com/text-mining/
3. Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS by Goutam Chakraborty, Murali Pagolu, Satish Garla
4. https://www.kaggle.com/rmalshe/comprehensive-analysis-of-the-phone-world

CONTACT INFORMATION

Name: Mohana Krishna Chaitanya Korlepara

Email: Chaitanya.korlepara@okstate.edu

Phone: 405-780-3782