SAS® GLOBAL FORUM 2018

USERS PROGRAM

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF

Which Smart Phone to Choose?

Krishna Chaitanya Korlepara, Student, Oklahoma State University

A Business Analytics student at Oklahoma State university with 4 years of experience in Business intelligence and Analytics across domains of marketing, retail and healthcare. Adept in working and coordinating with diverse teams to accomplish tasks. Successful in developing and implementing data driven business strategies with proven business and analytical skills.

@kmkchaitanya

Analysis of Unstructured Data: Topic Mining using Text

Data Discovery

Android smartphones have over 88% of the market share in the smartphone industry. With so many new phones launched with similar price point and features it almost makes 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. Buyer can narrow down their choice to a single product bases on their most desired aspect of a phone whereas manufactures 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 is to analyze the overall sentiment of the reviews of smartphones which fall in similar price point. In order to pursue this I have extracted the user comments from Amazon using a web crawler and user SAS enterprise miner and SAS sentimental
studio to draw insights.

Data Discovery

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 has over 400,000 customer reviews on various phones.

Data Discovery

- The data in the form of excel file is imported in SAS enterprise miner using file import node and analyzed using text parsing, text filter, text topic, text rule builder and text profile nodes
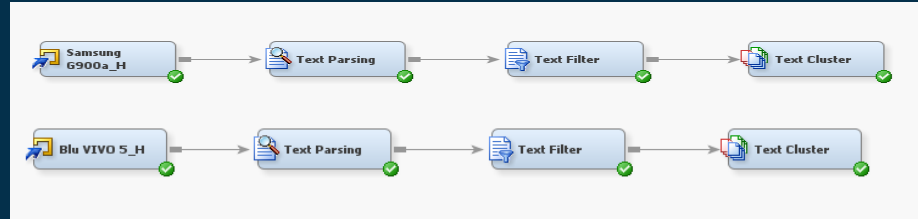
- Stratified Sampling is performed with Target as the Strata variable to segregate the dataset into Training (70%) and Validation (30%) datasets using the Data Partition node
- In order to process the text data, Text Parsing and Text Filter nodes are used. Text Filter Node applies filters to the text data and creates a transaction dataset that details which observations contain which words.

Term Weight analysis
- The Text parsing filter separates the reviews into words and ranks are assigned to each word according to the weight,phone,be,have all these have top ranks but we can filter the stop words using text filter.

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

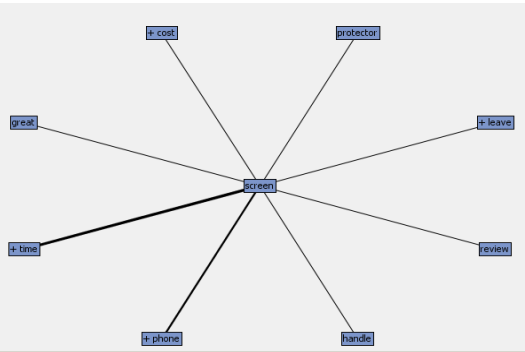# Analysis of Unstructured Data: Topic Mining using Text

Fig 1: Concept link negative review of 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 more during the usage which causes discomfort among the user which causes them to give poor review.
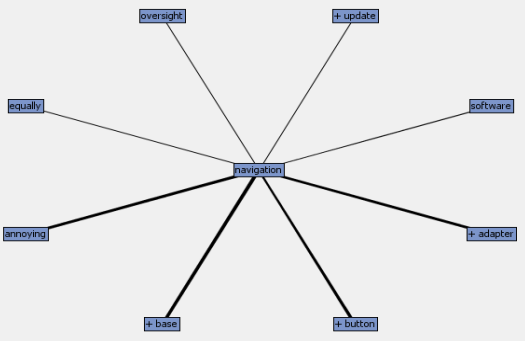

Fig 2:Concept link negative reviews of BLU VIVO5

From the reviews it can be seen that the navigation has high association between annoying which suggests that the phone's inbuilt software is causing problems to the users in navigation hence making them give negative reviews/comments about the phone.

- In the Samsung galaxy G900 the screen is strongly associated with time suggesting the response time of the screen is low
- Navigation in the smartphone BLU VIVO 5 is troublesome that is causing discomfort among the users.
- In the BLU VIVO 5 there are many common issues in front Camera.
- From the clusters we could find the good attributes that were mentioned in the reviews.

Discussion

This analysis was limited to only single of models of Samsung and Blu which 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 segmentation to give a good value for money factor for smartphones. 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.
80% of the real world data is in unstructured form this information can be leveraged to make improvements in the upcoming models and also 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

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

Results

- We could find few recurring issues that were causing customer dissatisfaction leading to a poor review.

Acknowledgment

#SASGF

SAS®
GLOBAL
FORUM
2018

April 8 – 11 | Denver, CO
Colorado Convention Center