

SAS® GLOBAL FORUM 2018

USERS PROGRAM

Analysis of Nokia Customer Tweets with SAS® Enterprise Miner™ and SAS® Sentiment Analysis Studio

Vaibhav Vanamala
MS in Business Analytics, Oklahoma State
University

April 8 - 11 | Denver, CO
#SASGF

Analysis of Nokia Customer Tweets with SAS® Enterprise Miner™ and SAS® Sentiment Analysis Studio

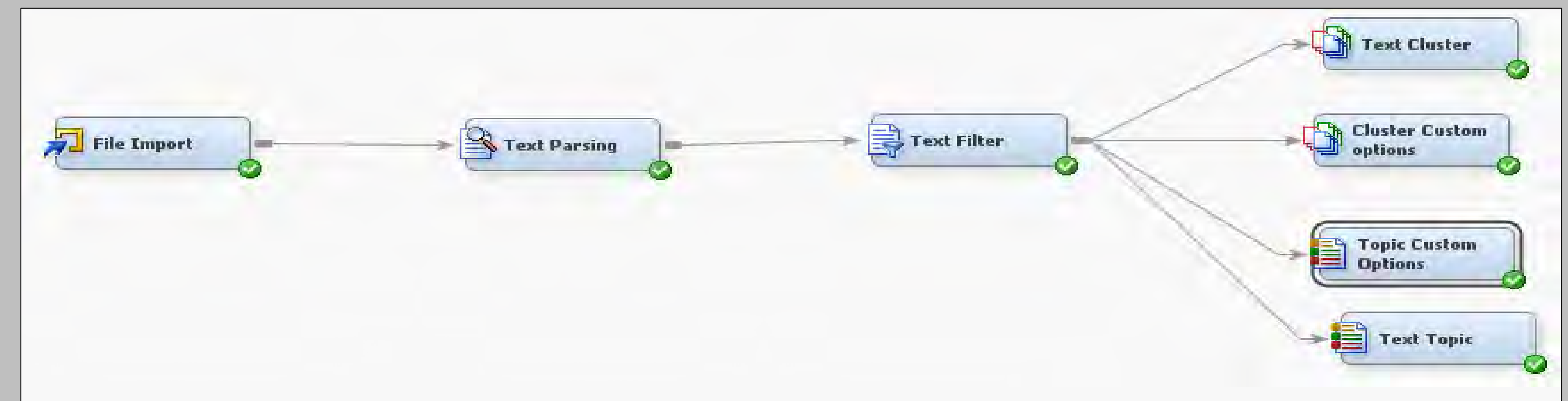
Vaibhav Vanamala

MS in Business Analytics, Oklahoma State University

ABSTRACT

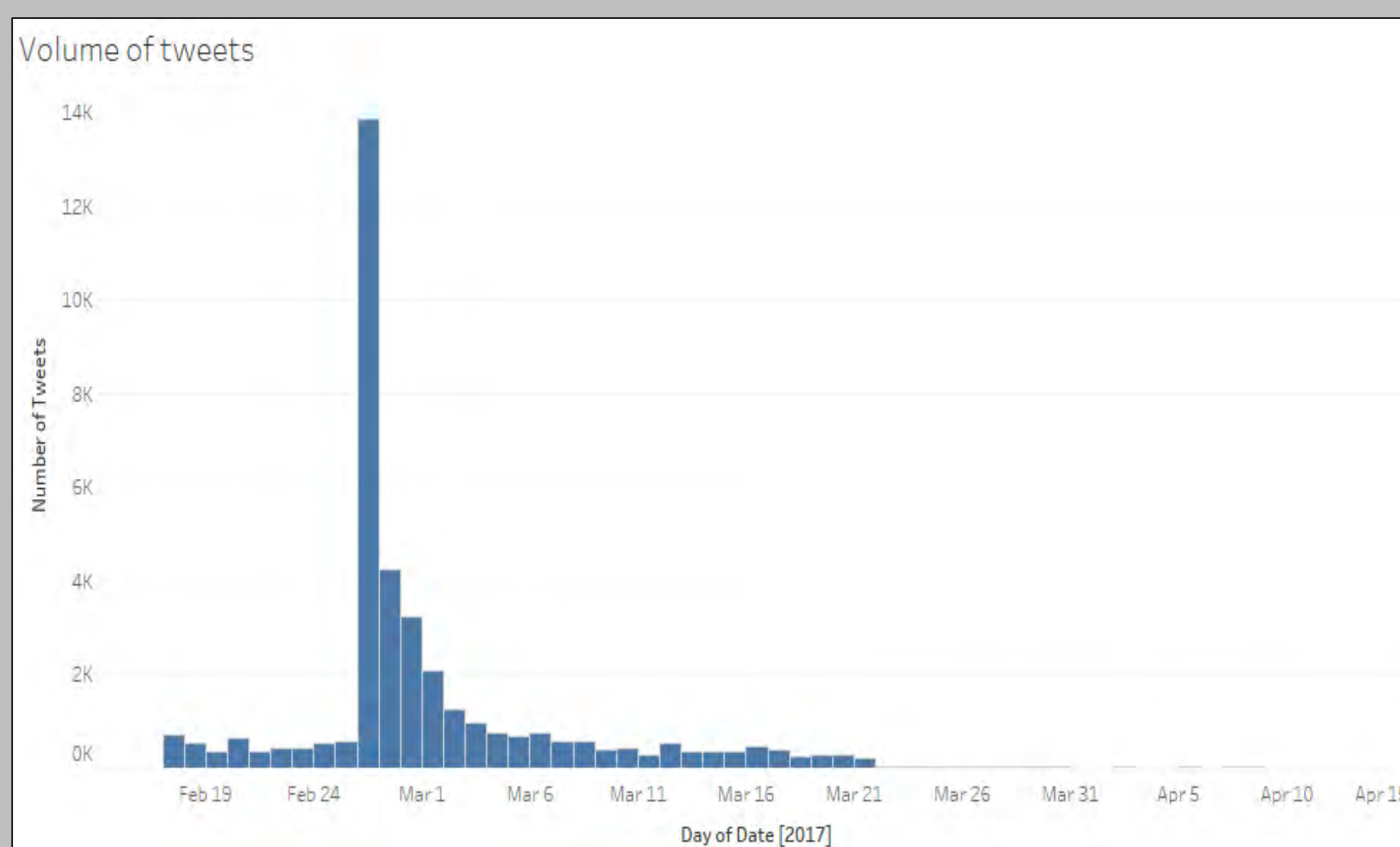
The launch of new Nokia phones has produced some significant and trending news throughout the globe. There has been a lot of hype and buzz going around the release of New Nokia phones in the mobile market at the MWC 2017 conference by HMD Technologies. Social media provides a platform for millions of people to share or express their opinions. There has been a significant magnitude of responses in the social media after the launch of the Nokia phones. In this paper, my goal is to analyze the overall sentiment prevailing in the social media posts. In order to achieve this, I have extracted real time data from twitter using google twitter API over a period of time and studied the responses of the people. I have used SAS® Enterprise Miner and SAS Sentimental Analysis Studio to evaluate key questions regarding the launch of Nokia phones such as understanding the needs and expectations of customers and perception of people about the launch of Nokia phones. In my results, NLP rule based model outperformed the default statistical model in SAS® Sentimental Analysis Studio for predicting sentiments in test data. NLP rule based model has provided deeper insights than statistical model in understanding consumers' sentiments.

METHODOLOGY



DATA PREPERATION

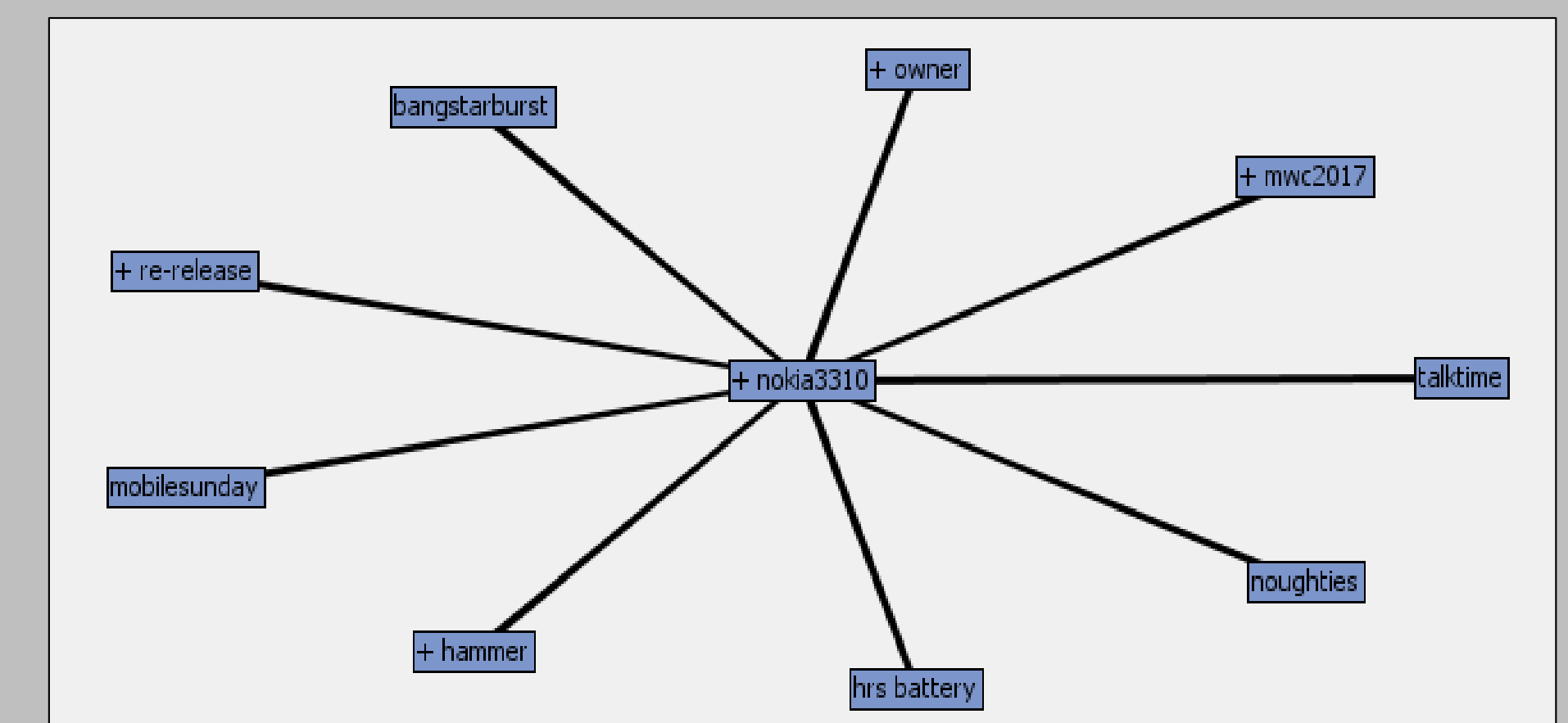
Collected tweets using twitter archiver over the span of 50 days (Feb 19th-April 17th, 2017). 40,000 tweets were collected for the Analysis



Variable Name	Type	Format	Length	Description
Date	Date	DATE9	8	Date on which the tweet was posted
Screen Name	String	CHAR20	20	Username of the twitter
Full Name	String	CHAR20	20	Name of the Tweeter
Tweet Text	String	CHAR 200	200	The actual text of the tweet
Tweet ID	String	CHAR20	20	Unique ID of the tweeter
App	String	CHAR50	50	Device through which the tweet was posted
Followers	Number	BEST12	8	Number of followers for that twitter
Follows	Number	BEST12	8	Number of fellow tweeters he/she follows
Retweets	Number	BEST12	8	Number of times hi/her tweet has been retweeted
Favourite	Number	BEST12	8	Total number of likes for that tweet
Verified	Number	Boolean	1	Flag of verification
User Since	DATE	DATE9	8	Time since the twitter account is active
Location	String	CHAR150	150	Location of tweet
Bio	String	CHAR150	150	Biography information the user

CONCEPT LINKS

- Long battery backup(hrs battery)
- Non shatterable strength of Nokia phones(hammer)
- Nostalgia effect of phones (noughties)
- New Nokia Owners (owner)
- Mobile World Congress(mwc2017)



Analysis of Nokia Customer Tweets with SAS® Enterprise Miner™

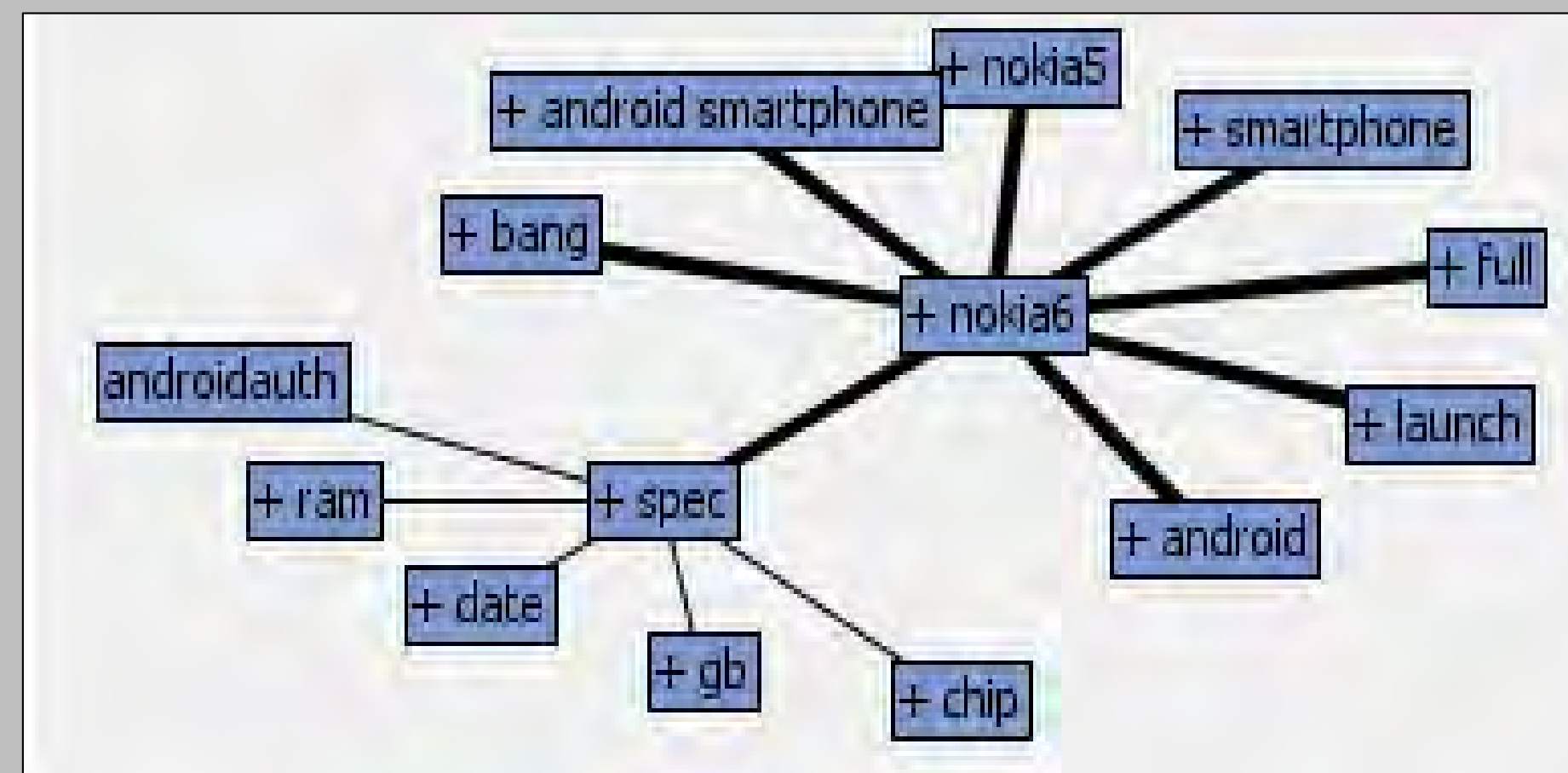
and SAS® Sentiment Analysis Studio

Vaibhav Vanamala

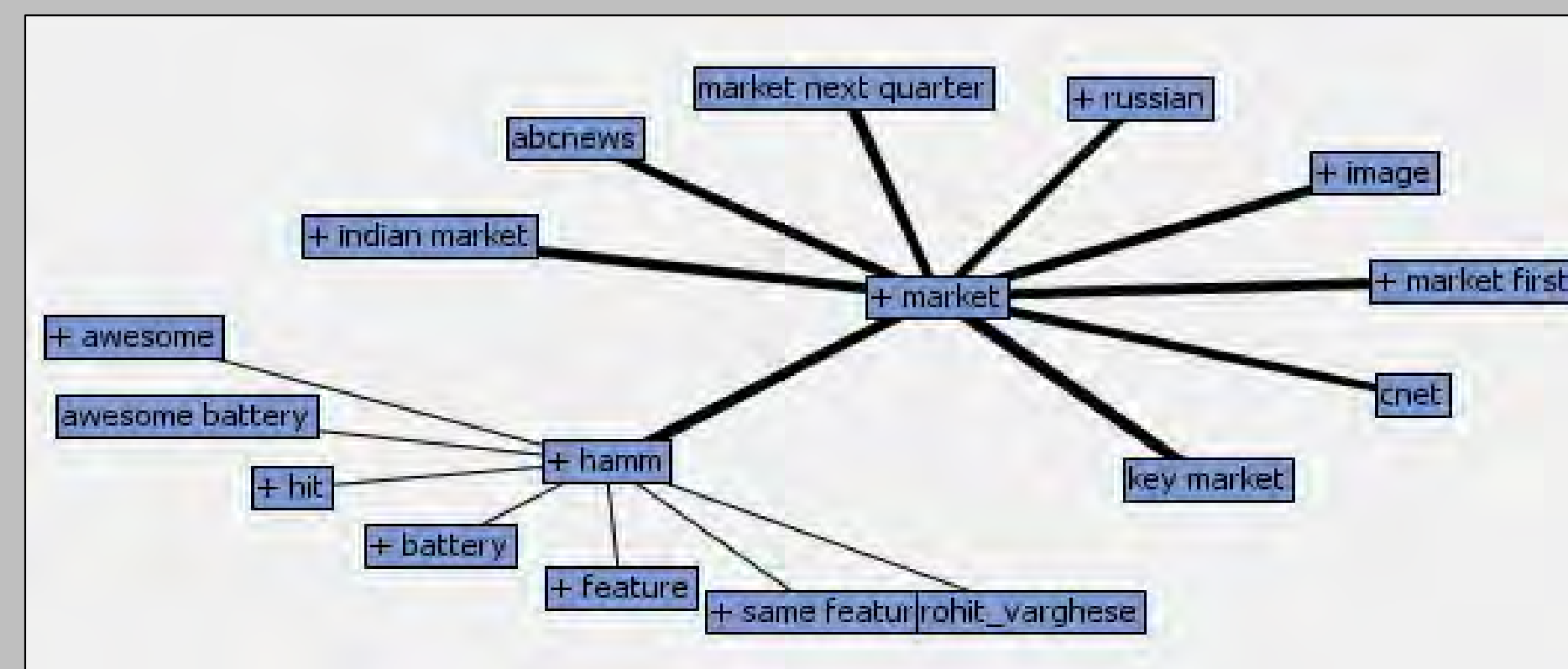
MS in Business Analytics, Oklahoma State University

CONCEPT-LINKS

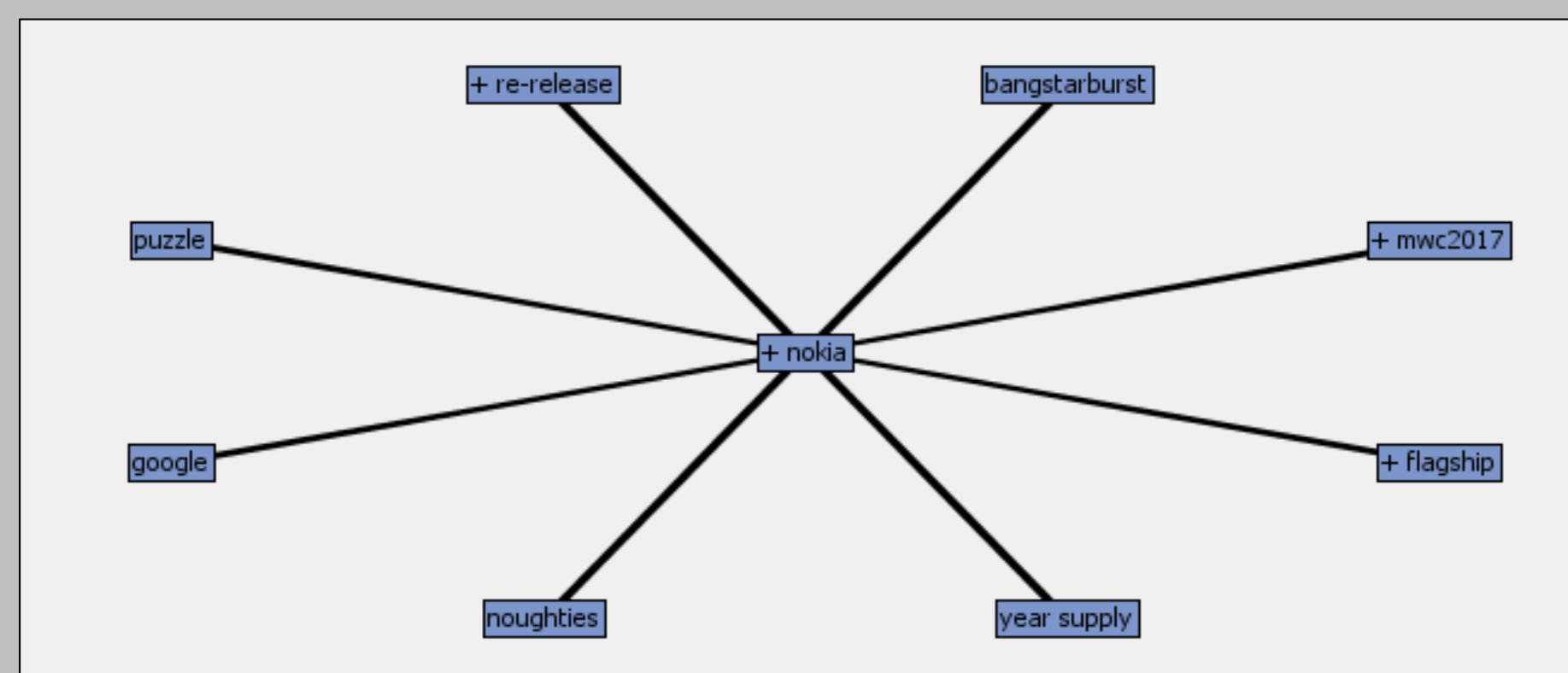
- Android (Nokia Mobile OS)
- Smartphone
- Specifications(Ram, Memory, Chip)
- Launch
- Nokia 5



- Indian Market
- Russian Market
- Next quarter of the year
- Market image
- Cnet , Abcnews

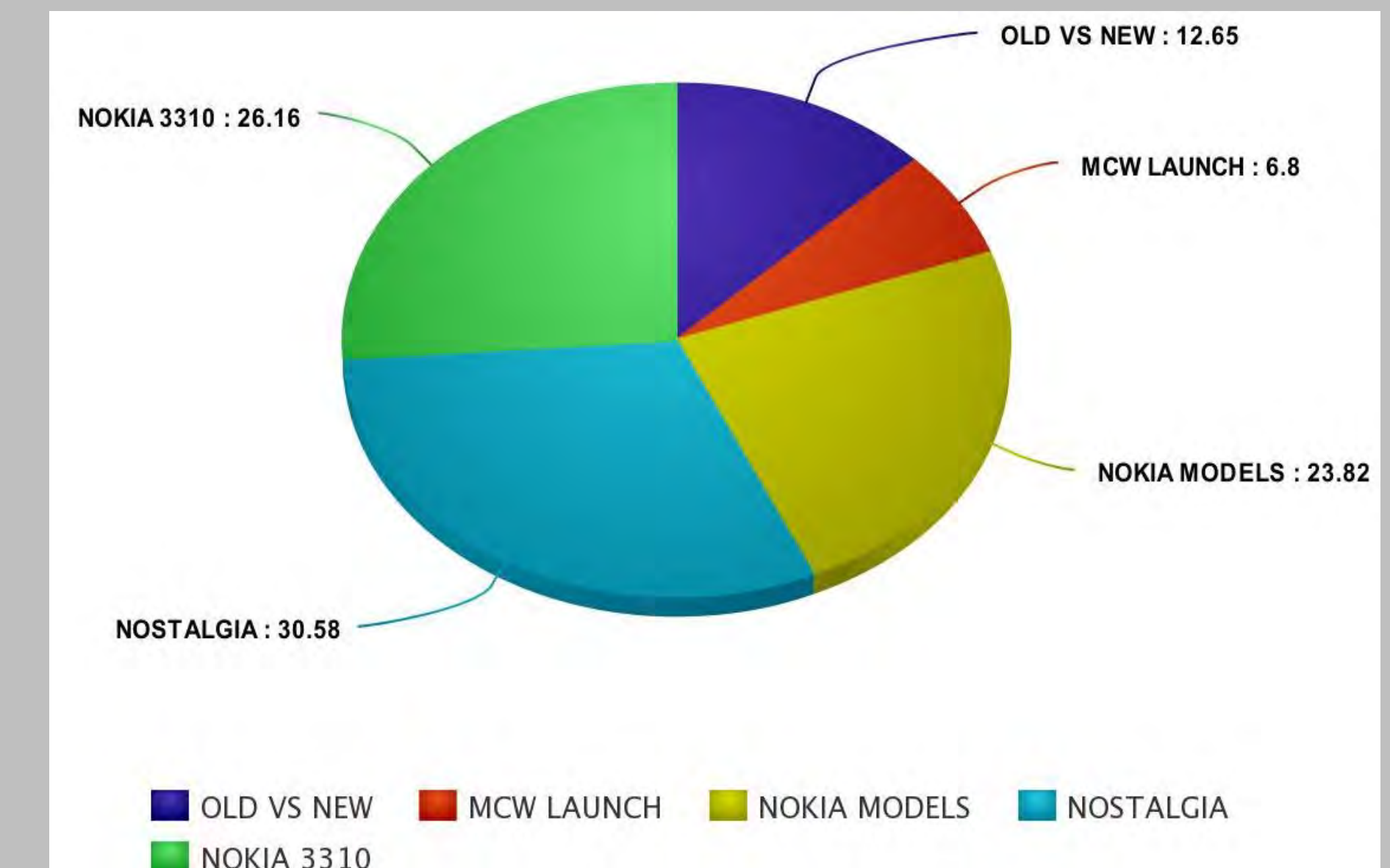


- Flashback coming into future (Noughties)
- Old Features like puzzles and Nokia ringtone
- Year long supply
- Flagship phone



CLUSTER ANALYSIS

1. OLD VS NEW
2. MCW LAUNCH
3. NOKIA MODELS
4. NOSTALGIA
5. NOKIA 3310



Cluster	Terms
Nokia Models	Nokia5 Nokia6 Nokia3 Nokia3310 Nostalgia Mcw2017 launch Comeback
Mcw Launch	MCW17 Smartphone Android Tech HMD Technology Market Release Features Good
Nostalgia	Snake game Fun Month-stand-by-time puzzle Hammer Nougat ball
Nokia 3310	Legend Snake Iconic Relaunch Bang start Return Comeback
Old vs New	Nokia 3310 old new comparision comeback strong

- Cluster 1 deals with the comparison of old phone and new phone.
- Cluster 2 mainly concentrates on MCW conference based on the launch of Nokia Phones.
- Cluster 3 is mainly concentrated with upcoming models of new Nokia phones.
- Cluster 4 focused about the Nostalgic features of Nokia3310 phone like Snake game, Hammer type resistance, Long battery life because of its features resembling its predecessor.
- Cluster 5 gives the perception of people by the re-launch of Nokia3310.

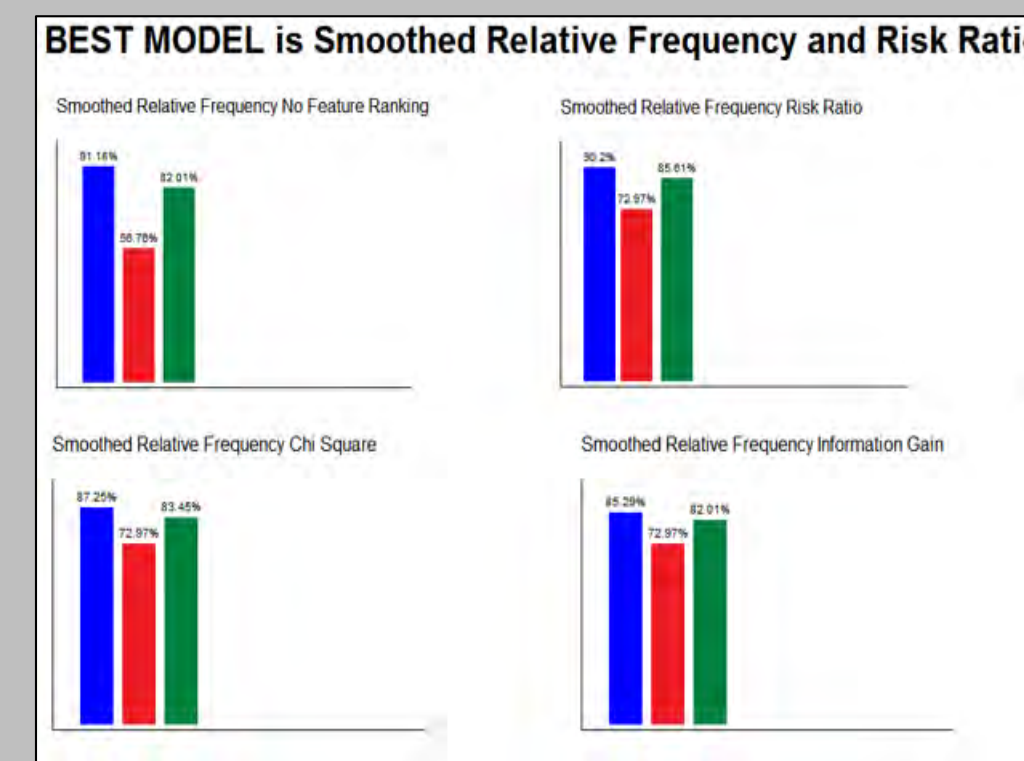
Analysis of Nokia Customer Tweets with SAS® Enterprise Miner™ and SAS® Sentiment Analysis Studio

Vaibhav Vanamala

MS in Business Analytics, Oklahoma State University

SENTIMENT ANALYSIS

Random sample of 800 tweets and coded them as either positive and negative. Neutral tweets were not considered. 408 positive tweets and 392 negative tweets. These 800 tweets were used to train the model.



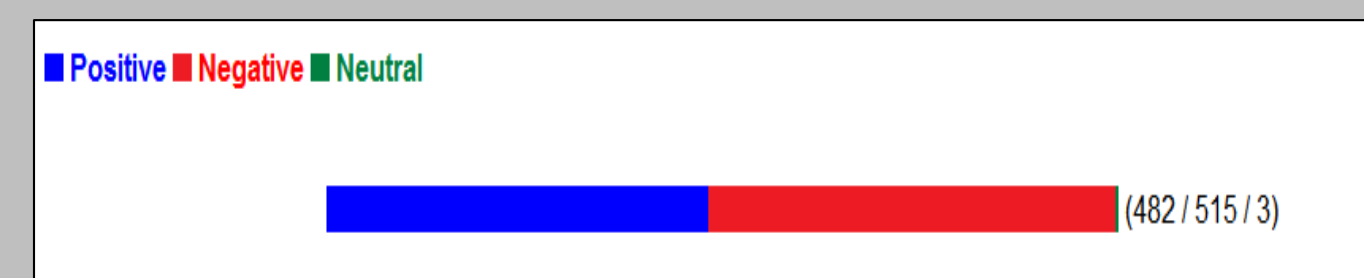
Statistical Model

The overall distribution of sentiments towards the comeback of Nokia were slightly negative (51.5%) and positive (48.2 %) and very less neutral tweets.



RULE BASED MODEL

The overall distribution of sentiments towards the comeback of Nokia were slightly negative (53.1%) and positive (47.9 %).



REFERENCES

Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS®

<http://nymag.com/selectall/2017/02/the-classic-nokia-3310-is-getting-an-update-and-coming-back.html>

<https://thenextweb.com/gadgets/2017/02/14>

INFERENCES

- Nokia has successfully created their brand awareness in the market
- Nokia's Marketing campaign of new Nokia phones has mixed opinions in the people
- People are skeptical about the performance of the phone compared to other phones
- Nokia has to enhance its features and also ensure better quality
- People are interested to see some more technological advancements in new Nokia phones

CONCLUSIONS

To conclude this paper, I have performed text mining and sentimental analysis of tweets regarding Comeback of Nokia phone after the launch of phone in MWC 2017 conference collected in the time frame between February to March. I have used Google Twitter API to download the tweets in real time and save them in spreadsheets online. I have used SAS Enterprise miner to clean and analyze the tweets. I have used the concept links to understand relationship between terms used in the tweets. I also came up with 5 clusters to understand the behavior of tweets. I then used SAS Sentimental Analysis studio to categorize sentiment of each tweet and then built a rule based model to predict the polarity of a tweet.

ACKNOWLEDGEMENT

I sincerely thank Dr. Goutam Chakraborty and Dr. Miriam McGaugh for their valuable guidance and motivation for accomplishing this paper.



SAS[®] GLOBAL FORUM 2018

April 8 - 11 | Denver, CO
Colorado Convention Center

#SASGF

Text and Sentiment Analysis of customer tweets of Nokia using SAS® Enterprise Miner™ and SAS® Sentiment Analysis Studio

Vaibhav Vanamala, Oklahoma State University, Stillwater, OK

ABSTRACT

The launch of new Nokia phones has produced some significant and trending news throughout the globe. There has been a lot of hype and buzz going around the release of New Nokia phones in the mobile market at the MWC 2017 conference by HMD Technologies. Social media provides a platform for millions of people to share or express their opinions. There has been a significant magnitude of responses in the social media after the launch of the Nokia phones. In this paper, my goal is to analyze the overall sentiment prevailing in the social media posts. In order to achieve this, I have extracted real time data from twitter using google twitter API over a period of time and studied the responses of the people. I have used SAS® Enterprise Miner and SAS Sentimental Analysis Studio to evaluate key questions regarding the launch of Nokia phones such as understanding the needs and expectations of customers and perception of people about the launch of Nokia phones. In my results, NLP rule based model outperformed the default statistical model in SAS® Sentimental Analysis Studio for predicting sentiments in test data. NLP rule based model has provided deeper insights than statistical model in understanding consumers' sentiments.

INTRODUCTION

Nokia used to be one of the world's biggest mobile phone manufacturers but it fell behind with the advent of iPhone and Android smartphones. The former giant couldn't catch up fast enough and quickly found itself bleeding and struggling to maintain its dominant market share. In 2014, Nokia's Mobile Devices and Services division was sold to Microsoft. In 2016, Finnish company HMD Global bought a part of Microsoft's feature phone business and has a licensing agreement that allows it to make smartphones under the Nokia brand. As a comeback in 2017, Nokia has launched a series of phones of different ranges into the market and knowing the perception, opinions and expectations of the its customers is very important as it helps in establishing and building their brand in the market.

Social media has evolved from a marketing channel to customer experience channel. It plays an important role in capturing people's sentiments and opinions about a wide range of products and services. Social media provides a platform which allows people to share or express their unbiased opinions. Attention to the opinions and feedback which customers provide about the products and services via social media is a critical factor to the success of the companies in the market place. In this paper, the social media that I chose is Twitter.

After the launch of Nokia phones, I have collected tweets from customers using twitter archiver in google spreadsheets. I have used these tweets to study the sentiment of the people regarding the Nokia phones in the market. The main objective of this analysis is to help Nokia manufacturers to improve the quality of the phones, to meet the expectations of the customers and to generate maximum revenue. This paper also focuses on understanding what customers liked and disliked, whether their reactions were either positive or negative about the newly launched products which may be one of the pioneer contributors to the revenue generated by the company.

TIMELINE

Timeline	Action Items	Timeframe
	Text Analysis and Sentimental analysis	Mar-Apr' 17
a.	Data Preperation and Cleaning	16th Mar - 18th Feb
b.	Descriptive Analysis	22nd Mar - 28th Mar
c.	Text Analysis to find out most frequent terms, concept links etc	1st Apr - 15th Apr
d.	Build a Statistical model combined with rule Based model to perform sentir	19th Apr- 27th Apr

Figure 1. Timeline of the project

DATA

I first extracted data from Twitter by using Twitter archiver add-on in Google Spreadsheets. Twitter archiver in Google Spreadsheet runs continuously and stores data on the cloud automatically which ensures that there is no loss of a single tweet for a given hashtag. I have chosen a time period of 50 days starting from the announcement of Nokia Phones at MWC conference (Feb 19th,2017- April 17th, 2017).

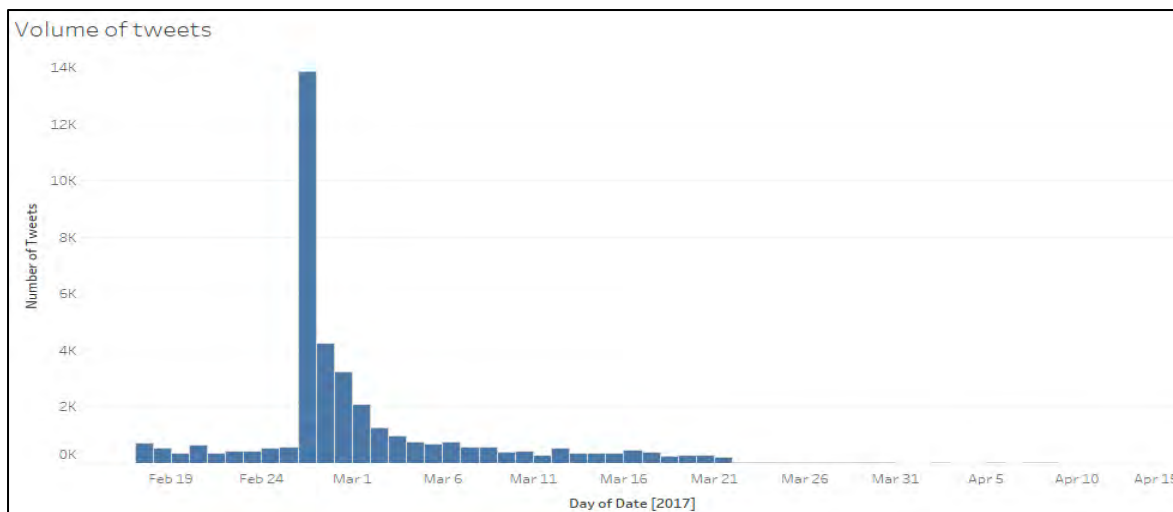


Figure 2. Volume of tweets collected over the time

A corpus of around 40,000 tweets in English language were published by Twitter users in the chosen time period.

Metadata of the datasets used is as shown below.

Variable Name	Type	Format	Length	Description
Date	Date	DATE9	8	Date on which the tweet was posted
Screen Name	String	CHAR20	20	Username of the twitter
Full Name	String	CHAR20	20	Name of the Tweeter
Tweet Text	String	CHAR 200	200	The actual text of the tweet
Tweet ID	String	CHAR20	20	Unique ID of the tweeter
App	String	CHAR50	50	Device through which the tweet was posted
Followers	Number	BEST12	8	Number of followers for that twitter
Follows	Number	BEST12	8	Number of fellow tweeters he/she follows
Retweets	Number	BEST12	8	Number of times hi/her tweet has been retweeted
Favourite	Number	BEST12	8	Total number of likes for that tweet
Verified	Number	Boolean	1	Flag of verification
User Since	DATE	DATE9	8	Time since the twitter account is active
Location	String	CHAR150	150	Location of tweet
Bio	String	CHAR150	150	Biography information the user

Figure 3. Metadata of the dataset

DATA PROCESSING USING SAS ENTERPRISE MINER:

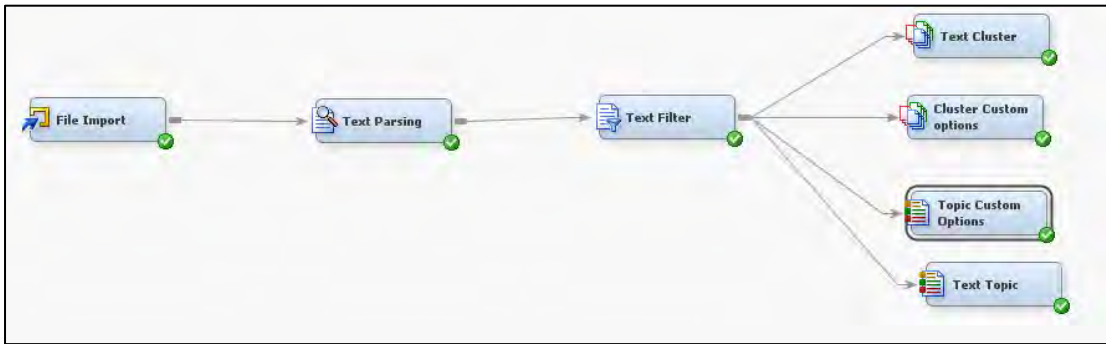


Figure 4. ER Diagram of the project

For the analysis, I have considered all the tweets and followed the general text analytic approach suggested by Chakraborty, Pagolu and Garla (2013). This involved using concepts like NLP techniques, lemmatization, concept linking, use of synonyms, etc.

Concept links help in understanding the relationship between words (terms) based on the co-occurrence of words (terms) in the document. It shows the terms that are strongly associated with the given term. Thick links indicate strong association between the terms.

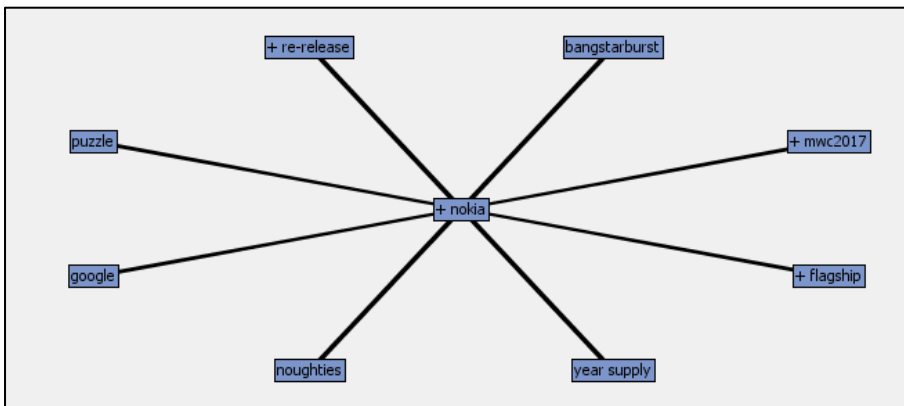


Figure 5. Concept link of the word "NOKIA"

From Figure 5, we can see that people are considering Nokia as their flashback coming into future (Noughties) and are mostly interested to find features like puzzles (snake game, Nokia ringtone). People are excited about the re- release of the phone in the MWC2017 conference and are waiting to see yearlong supply of the flagship phone.

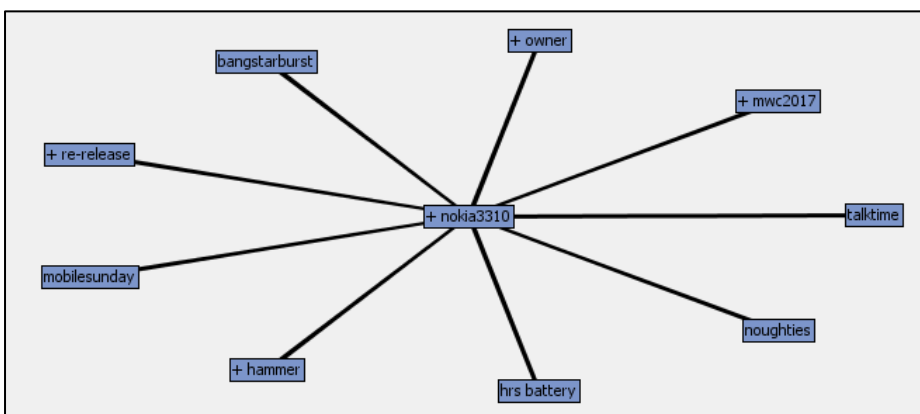


Figure 6. Concept link of the word "NOKIA3310"

From Figure 6, we can say that most people are interested in the specifications of the nokia3310 mobile like long battery backup, non-shatter able strength of Nokia phones(hammer), and also the nostalgia effect of phones (noughties) and also people are much interested in the new owners who took over the new Nokia.

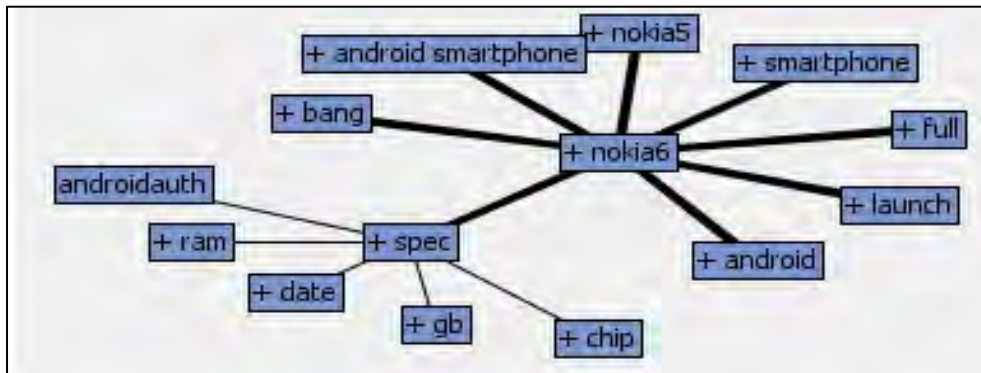


Figure 7. Concept link of the word “NOKIA6”

From Figure 7, we can see the features and specifications of the Nokia 6 that people are interested.

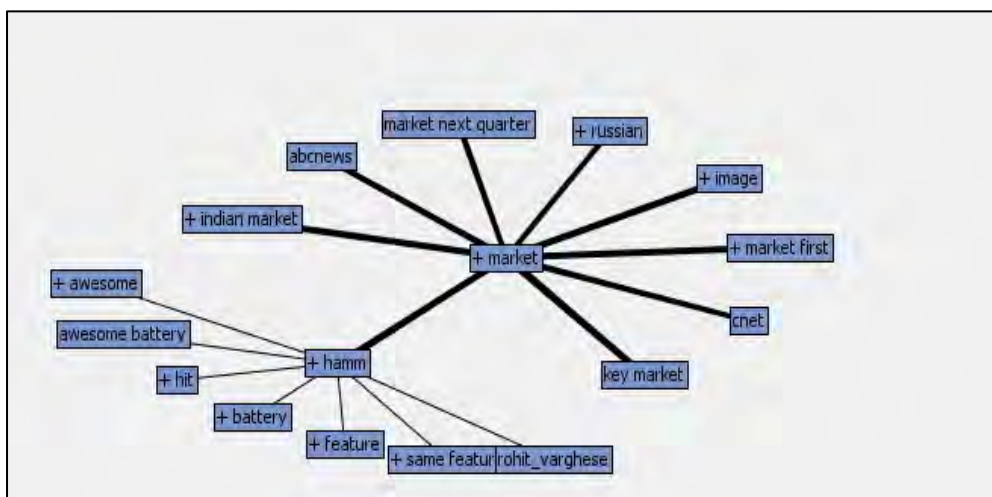


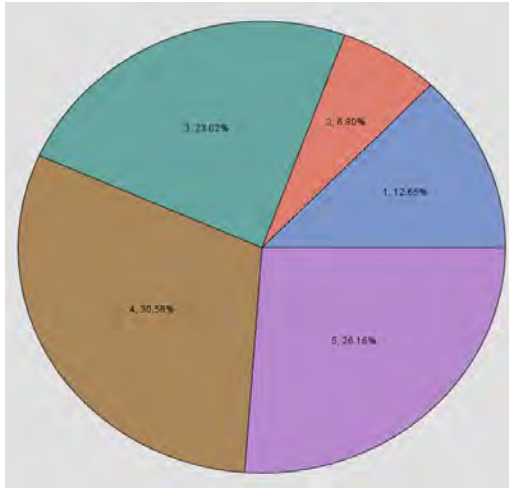
Figure 8. Concept link of the word “MARKET”

From Figure 8, people from Indian market and Russian are more interested in buying the new Nokia phone and the phone is going to release in next quarter of the year. The new company is building its Market image based on Nokia Nostalgic fans.

Finding Patterns in the Tweets

Cluster	Terms
Nokia Models	Nokia5 Nokia6 Nokia3 Nokia3310 Nostalgia Mcw2017 launch Comeback
Mcw Launch	MCW17 Smartphone Android Tech HMD Technology Market Release Features Good
Nostalgia	Snake game Fun Month-stand-by-time puzzle Hammer Nougat ball
Nokia 3310	Legend Snake Iconic Relaunch Bang start Return Comeback
Old vs New	Nokia 3310 old new comparision comeback strong

Figure 9. Clusters found in the data



- I. Old Vs New
- II. MCW Launch
- III. Nokia Models
- IV. Nostalgia
- V. Nokia 3310

Figure 10. Percentage of Clusters present in the data

The above clusters provide us a brief idea about the perception of people on the re-launch of Nokia phone. The tweets from people are categorized into these 6 clusters and the following characteristics can be observed from these clusters

- One of the Cluster is focused about the Nostalgic features of Nokia3310 phone like Snake game, Hammer type resistance, Long battery life because of its features resembling its predecessor.
- One Cluster is mainly concentrated with upcoming models of new Nokia phones.
- One cluster gives the perception of people by the re-launch of Nokia3310.
- Another Cluster mainly concentrates on MCW conference based on the launch of Nokia Phones.
- Another Cluster deals with the comparison of old phone and new phone.

SAS SENTIMENTAL ANALYSIS ON TWITTER TWEETS:

Based on initial Analysis of tweets, I figured out that there was lot of excitement and anticipation in the tweets for the launch of new Nokia phones which showed good manifestation of positivity and negativity and they are sufficient to perform sentimental Analysis.

For this, I have taken a random sample of tweets around 800 from the entire dataset I read through each of the tweets diligently, and coded them as either positive and negative. I did not take the neutral tweets into consideration.

Overall, there are 408 positive tweets and 392 negative tweets in the dataset used for modelling.

After performing the above steps, positive and negative tweets were linked to the model in SAS Sentimental Analysis studio. The model was trained using the segmented tweets to produce the following results for a statistical model. Smoothed Relative Frequency and Risk Ratio model was selected as best model.

STATISTICAL MODEL

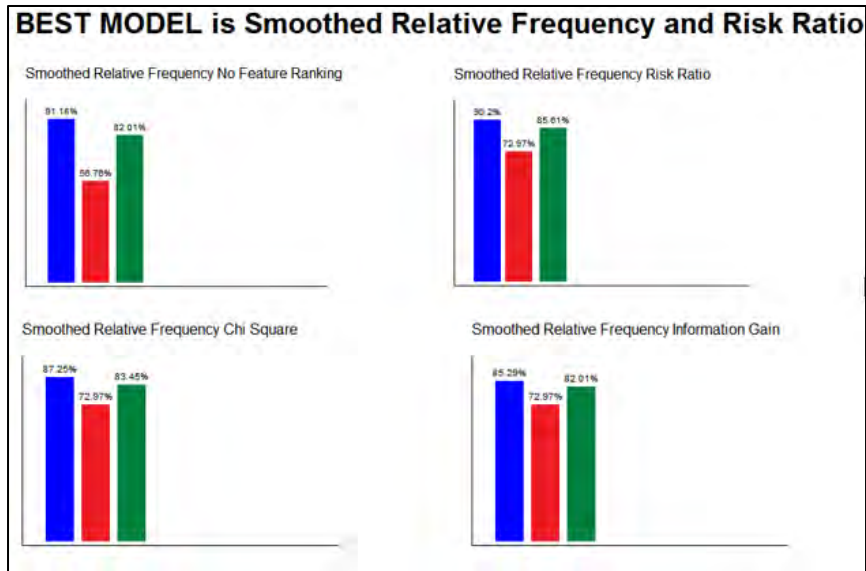


Figure 11. Statistical Model

I then took a small dataset of 1000 tweets that replicates the original dataset to score using the model and The overall distribution of sentiments towards the comeback of Nokia were slightly negative (51.5%) and positive (48.2 %) and very less neutral tweets.

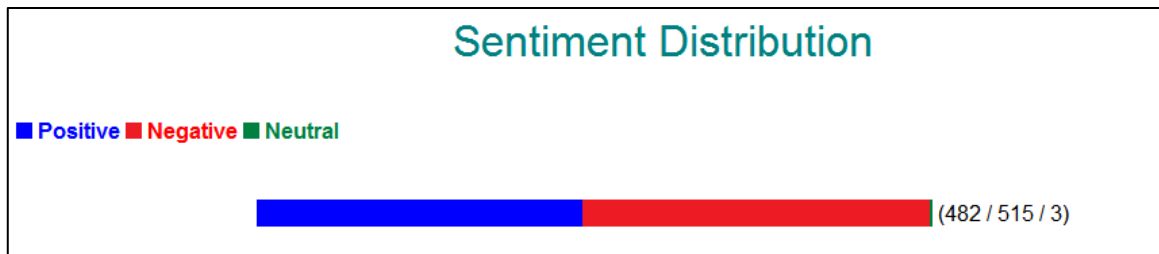


Figure 12. Sentiment Distribution of the rule based model

RULE BASED MODEL

In order to build a Rule based model. I have imported around 100 positive rules and 100 negative rules from the learned features of the statistical model that was built previously. Rules which were not appropriate were removed. The rules which were omitted contained unrelated and neutral words.



Figure 12. Sentiment Distribution of the rule based model

The overall distribution of sentiments towards the comeback of Nokia were slightly negative (53.1%) and positive (47.9 %).

INFERENCES

- Nokia operated by the new owners has successfully created their brand awareness in the market and are one of the strong contenders to share the mobile market in future.
- Nokia's Targeted Marketing campaign based on Nostalgic features of new Nokia phones has mixed opinions in the people. Even though they are excited about the launch of the product, they are sceptical about the performance of the phone compared to other phones.
- People who belong to 20th century are more inclined towards buying Nokia phone, but people in the 21st century are less likely towards buying Nokia Phone.
- People are looking at the Nokia phones as a stand by phone rather than Smart phone. People are interested to see some more technological advancements in new Nokia phones.
- Nokia has to enhance its features and also ensure better quality to attract customers.

CONCLUSIONS:

To conclude this paper, I have performed text mining and sentimental analysis of tweets regarding Comeback of Nokia phone after the launch of phone in MWC 2017 conference collected in the time frame between February to March. I have used Google Twitter API to download the tweets in real time and save them in spreadsheets online. I have used SAS Enterprise miner to clean and analyze the tweets. I have used the concept links to understand relationship between terms used in the tweets. I also came up with 5 clusters to understand the behavior of tweets. I then used SAS Sentimental Analysis studio to categorize sentiment of each tweet and then built a rule based model to predict the polarity of a tweet.

ADDITIONAL WORK:

Planned to improve the efficiency of predicting polarity of sentiment by incorporating emoticons in sentimental mining by developing a macro that cleans the tweets by replacing with equivalent text.

REFERENCES

- <https://support.sas.com/resources/papers/proceedings15/3330-2015.pdf>
- <http://nymag.com/selectall/2017/02/the-classic-nokia-3310-is-getting-an-update-and-coming-back.html>
- http://www.huffingtonpost.com/entry/nokia-phone-comeback_us_58a4701de4b03df370dc30be.
- <http://www.androidpolice.com/2017/03/01/nokias-comeback-explained-historic-phone-company-suddenly-making-android-phones/>
- https://thenextweb.com/gadgets/2017/02/14/no-seriously-the-nokia-3310-is-coming-back/#.tnw_IzC7MYeB

ACKNOWLEDGEMENTS

I sincerely thank Dr. Goutam Chakraborty for his valuable guidance and motivation for accomplishing this paper.

I also thank Dr. Miriam McGaugh for her constant support and suggestions throughout this study.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Vaibhav Vanamala

Oklahoma State University

Phone: +1(405)-612-2773

Email ID: VAIBHAV.VANAMALA@OKSTATE.EDU

LinkedIn: <https://in.linkedin.com/in/vaibhav-vanamala>

Skype: Vaibhav. Vanamala

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.