SCSUG 2020 - Paper TT03

Sentiment Analysis on P&G products' customer reviews using SAS Enterprise Miner and SAS Sentiment Analysis Studio

Trinh Phan, Oklahoma State University, Stillwater, OK

Mentor: Reshma Narayana, General Mills, MN

ABSTRACT

Consumer goods industry has seen tremendous growth in digital technology where online shopping has become a common habit among millions of people throughout the world. Statista forecasts that the Home and Laundry Care market will produce 6% of total revenue through online sales by 2023. Moreover, the recent Nielsen's consumer report has identified some significant changes in consumer behavior in the light of the Corona virus outbreak. Rise in online sales, drop in in-store shopping and increased demand of cleaning products. Therefore, online customer reviews continue to be an essential source of information for enhancing customer experience and business operation in many organizations. However, most of those comments comprise of unstructured data, making it unfeasible to perform traditional analysis. In addition, it would be inefficient for potential buyers to read all the reviews to decide on the best product out of a large pool of similar products. This project aims to apply opinion mining on data collected on P&G's reviews to identify main features and determine their relationship with customer satisfaction or dissatisfaction. Analysis of the textual data using text parsing, text topic, text cluster for identifying features and classification of sentimental reviews using predictive models (Text Rule Builder, Logistic Regression, Decision Tree, Neural Network) are deployed within SAS Text Miner 14.3. SAS Sentiment Analysis Studio is used to identify positive and negative reviews from all comments. Our findings suggest that cleaning power, fresh scent and great value have highly affected the customer satisfaction, whilst negative experience with packaging design for delivery and chemical smell were associated with customer dissatisfaction. Among predictive models, Text Rule Builder performs the best with a validation misclassification rate of 10.9% and average square error of 2.46%. Thus, the results of this paper could potentially help organizations in FMCG to enhance their product quality and improve operations for increasing customer experience.

INTRODUCTION

Established in 1837, The Procter & Gamble Company (P&G) has grown from a small family–run business to be one of the largest multinational consumer goods corporation. It offers an enormous variety of personal and health care, and hygiene products. According to its annual report in 2019, Fabric and Home Care generated the highest percentage of Net Sales (33%) and highest percentage of Net Earnings (19%) among reportable segments. This segment includes a wide range of fabric care products, including laundry detergents, additives and fabric enhancers; and home care products, including dishwashing liquids and detergents, surface cleaners and air fresheners. As a global leader in Fabric and Home Care, P&G is looking for an effective way to improve customer satisfaction. Besides data mining with numeric predictors, customer feedbacks enables company to capture the accurate and complete picture of customer experience. With the fast evolution of e-commerce platforms, people easily share their thoughts about products and services, criticizing or complimenting them. This analysis will examine around forty thousand reviews of thirty brands in the Fabric and Home Care segment, which have ratings ranging from 2.6 to 5. The result would help us understand how people think about those products and how data-mining technique is utilized to do sentiment in the new comments.

DATA PREPARATION

This study collected consumer reviews for P&G's Home care and Farbic care product from different websites such as Amazon and Home Depot in three markets UK, US and CA. The dataset is available on Kaggle in 2019 and stored in '.csv' (comma seperated values) files. The train data includes 42,200 reviews and 9 variables where each row indicates a specific review. Another dataset, the test data comprises 14,063 observations, which would be scored based on the best predictive model. The sentiment variable is created

as target variable for data mining model. This variable contains two levels as postive (rating >3) and negative (<=3). The train data description is as follow:

Variable	Type	Description
ID	Char	A unique identification of review
category	Char	Main categories are Homecare and Fabric Care
brand	Char	Brand name in each category
sub_category	Char	Subcategory in each brand
product_description	Text	Product name and quantity
market	Char	Country of reviewers: UK, US, CA
review_title	Text	Title of the review
review_text	Text	The content of review
review_rating	Num	5-star satisfaction rating
sentiment	Char	Target variable whether positive or negative

Table 1: Metadata

The below figure shows the sample extracted data:

id	category	brand	sub_category	product_description	market	review_title	review_text	review_rating
Te-1	Fabric Care	Tide	Laundry	Tide Pods He Turbo Laundr	US	The best in a very c	The best general wash detergent. Convenient contai	5
Te-2	Fabric Care	Tide	Laundry	Tide Washing Machine Cle	US	First time	This helped to clean our washing machine after getti	5
Te-3	Home Care	Cascade	Auto Dishwashing	Cascade Platinum ActionPa	US	I've been using ano	I've been using another well known brand and didn't	5
Te-4	Fabric Care	Tide	Laundry	Tide Purclean Liquid Laund	US	Great laundry deter	Smell is great and clothes are always clean. Great pr	5
Te-5	Home Care	Mr Clean	Surface Care	Mr. Clean Magic Eraser Cle	US	Five Stars	Good product, works well.	5
Te-6	Home Care	Cascade	Auto Dishwashing	Cascade Fresh Scent Dishw	US	Very Convenient	I get these ActionPacs monthly through subscribe an	5
Te-7	Home Care	Swiffer	Surface Care	180 Dusters Multi Surface	US	Greater helper	Hint: under furniture, under bed, spray a product like	5
Te-8	Home Care	Cascade	Auto Dishwashing	Cascade Platinum Actionpa	US	Five Stars	Best one for your dish washer!	5
Te-9	Home Care	Cascade	Auto Dishwashing	Cascade Dishwasher Deter	US	Platinum pacs	At 76 years old, can not believe this product (it is am	5
Te-10	Fabric Care	Tide	Laundry	Tide Original Scent Liquid L	US	So in love!	This is the only laundry detergent I use.It doesn't ma	5
Te-11	Fabric Care	Tide	Laundry	Tide PODS Plus Downy HE	US	Easy to use and sm	This is our new favorite type of laundry soap. It leav	5
Te-12	Home Care	Febreze	Air Care	Febreze Air Effects Heavy D	US	The best	Great scent, covers all orders, will only buy this versi	5
Te-13	Home Care	Febreze	Air Care	Febreze Air Effects Heavy D	US	Heavy duty smell el	Fantastic for clearing odors from the home; especia	5
Te-14	Home Care	Cascade	Auto Dishwashing	Cascade Complete Gel Citr	US	This product leaves	My dishes come out sparkling clean. I live in an area	5
Te-15	Home Care	Cascade	Auto Dishwashing	Cascade Platinum ActionPa	US	Best Ever!	Best dishwashing detergent I've ever used. They cost	5
Te-16	Home Care	Swiffer	Surface Care	Swiffer Wetjet Refill, Origi	US	Very good.	These are very good. They pick up most accumulatic	4
Te-17	Fabric Care	Dreft	Fabric Enhancer	Dreft Blissfuls In-Wash Sce	US	One Star	Too much scent for my newborn to be around. I wo	1
Te-18	Home Care	Cascade	Auto Dishwashing	Cascade Dishwasher Deter	US	Squeeky Clean Dish	We have hard water which makes getting clean dish	5
Te-19	Home Care	Dawn	Hand Dishwashing	Dawn Ultra Dishwashing Li	US	its the best	its dawn. its the best ha	5
Te-20	Home Care	Febreze	Air Care	Febreze Crisp Clean Air Fre	US	Five Stars	Smells very nice.	5
Te-21	Home Care	Swiffer	Surface Care	Swiffer Sweep+Vac Cordle	US	Cleaning tool	Works perfectly, great tool for quick cleaning	5
Te-22	Fabric Care	Gain	Fabric Enhancer	Gain Fireworks In-Wash Or	US	Five Stars	My favorite scent of all the brands.	5
Te-23	Home Care	Cascade	Auto Dishwashing	Cascade Complete Dawn D	US	This product cleans	I feel this product works a lot better than the gel poc	5
Te-24	Home Care	Cascade	Auto Dishwashing	Cascade AP Fresh Scent 60	US	easy to use and gre	cascade always has been my go to dishwashing dete	5
Te-25	Fabric Care	Tide	Laundry	TIDE LQ PODS SM 4/81 CT	US	I guess the washers	I guess the washers in my building can't handle these	3
Te-26	Fabric Care	Gain	Fabric Enhancer	Gain Fireworks In-Wash Or	US	Smells great! But H	Smells great! But HATE the pods. We got the dash bu	1

Figure 1. Sample of train dataset

DATA ANALYSIS

SAS Text miner has been employed to perform text analysis and extract valuable patterns about positive and negative reviews. This paper comprises of 3 parts:

1. Analyze the reviews for all brands to understand what people say about the product and compare some predictive models with Text rules builder.
2. Analyze positive and negative comment separately.
3. Do sentiment analysis on top 2 products representing each category: Tide(Fabric Care) and Cascade (Home Care)

ANALYSIS 1 – REVIEW ALL COMMENTS

The project involves a series of text analysis nodes: Text Parsing, Text Filter, Text Topic, Text Cluster. Then the textual data is input to different predictive models as shown below. The given workflow is set based on recommended practices in text analytics (Chakraborty, G., Pagolu, M., & Garla, S. 2014).

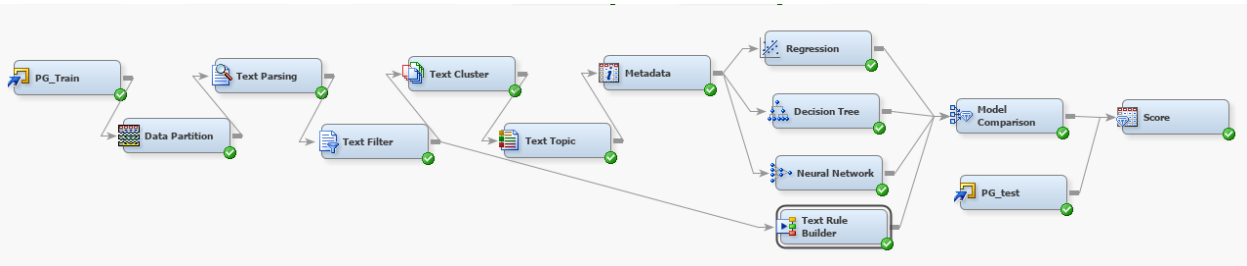


Figure 2. Workflow diagram

DATA PARTITION

The File import node is used to import the train dataset and then divided into two parts in SAS Enterprise Miner. The model will be trained based on 80% reviews and 20% was used for validation.

TEXT PARSING

The Text Parsing node is used to split a sentence into separate terms and properties are set as follows:

- "Detect Different Parts of Speech" is set as No to treat similar words as same.
- "Detect Find Entities" is set to standard.
- "Ignore Parts of Speech" includes these choices: "Abbr", "Aux", "Conj", "Det", "Interj","Part","Prep","Pron","Num" and "Prop".
- "Ignore Types of Attribute" contains "Num" and "Punct".

Term	Role	Attribute	Freq	# Docs	Keep ▼	Parent/Child Status	Parent ID	Rank for Variable numdocs
+ clean		Alpha	11938	9339	Y	+	7481	3
+ smell		Alpha	11803	8851	Y	+	31356	5
+ good		Alpha	9038	7799	Y	+	9056	7
+ great		Alpha	8369	7348	Y	+	23521	9
+ product		Alpha	9269	7344	Y	+	17901	10
+ love		Alpha	7470	6573	Y	+	16280	11
+ review		Alpha	5714	5641	Y	+	14587	12
+ collect		Alpha	5284	5281	Y	+	17416	14
+ promotion		Alpha	5244	5244	Y	+	13816	15
this review	Noun Group	Alpha	5240	5240	Y		7843	16
+ work		Alpha	5747	5194	Y	+	15527	17
clothes		Alpha	5753	4332	Y		33036	19
+ dish		Alpha	5352	4123	Y	+	8368	20
+ scent		Alpha	5594	4106	Y	+	37442	21
+ wash		Alpha	3976	3079	Y	+	3314	26
+ buy		Alpha	3494	3078	Y	+	18276	27
+ leave		Alpha	3221	2943	Y	+	761	29
well		Alpha	3066	2888	Y		9189	30
+ detergent		Alpha	3671	2878	Y	+	32024	31
+ fresh		Alpha	3041	2649	Y	+	33619	35
+ dishwasher		Alpha	3312	2487	Y	+	36257	37
always		Alpha	2650	2370	Y		23294	38
+ laundry		Alpha	2999	2360	Y	+	16284	39
+ price		Alpha	2477	2324	Y	+	34597	40

Figure 3. Text parsing result for train data

According to text parsing result, the most frequent terms are clean, smell, great, love, clothes, dish, wash, price. Those terms are related to home care and fabric care product. Moreover, it seems that product quality and value were primary causes for customer satisfaction.

TEXT FILTER

The output from the Text parsing node is passed to the Text Filter to show the most frequent or least important words. This step helps to eliminate unecessary information for further analysis. The properties are set as follows:

- "Check spelling" is set to Yes to correct the mispelled words.
- "Term weight" is set to mutual information.
- "Minimum number of document" is set to 3.

	TERM	FREQ	# DOCS	KEEP ▼	WEIGHT	ROLE	ATTRIBUTE
⊞	easy	2537	2330	☑	0.039		Alpha
⊟	price	2481	2328	☑	0.024		Alpha
	princess	1	1				Alpha
	prices	59	59				Alpha
	pricing	30	30				Alpha
	price	2291	2171				Alpha
	priced	97	97				Alpha
	priceï	2	2				Alpha
	price	1	1			Miscellaneous Pr...	Entity
⊞	time	2399	2090	☑	0.167		Alpha
⊞	pod	3038	2060	☑	0.235		Alpha
⊞	last	2205	2015	☑	0.069		Alpha
⊞	cascade	2643	1966	☑	0.044	Location	Entity
⊞	little	1976	1773	☑	0.025		Alpha

Figure 4. Text Filter results and Misspelled Words

Interactive filter viewer enables us to see which words are kept or dropped based on the weight and how the software processes the misspelled words, merging similar words into a group.

CONCEPT LINK

From the Interactive Filter Viewer of Text Filter properties panel, we can see the association of a particular term. The word of interest is at the center and is surrounded with other words, which are associated with that word. Concept Link is only available for a single term at a time. Therefore, we will explore some important words from the context. The width of the line indicates the strength of association, the thicker line, the higher the association. Some high frequency words are examined as follows:

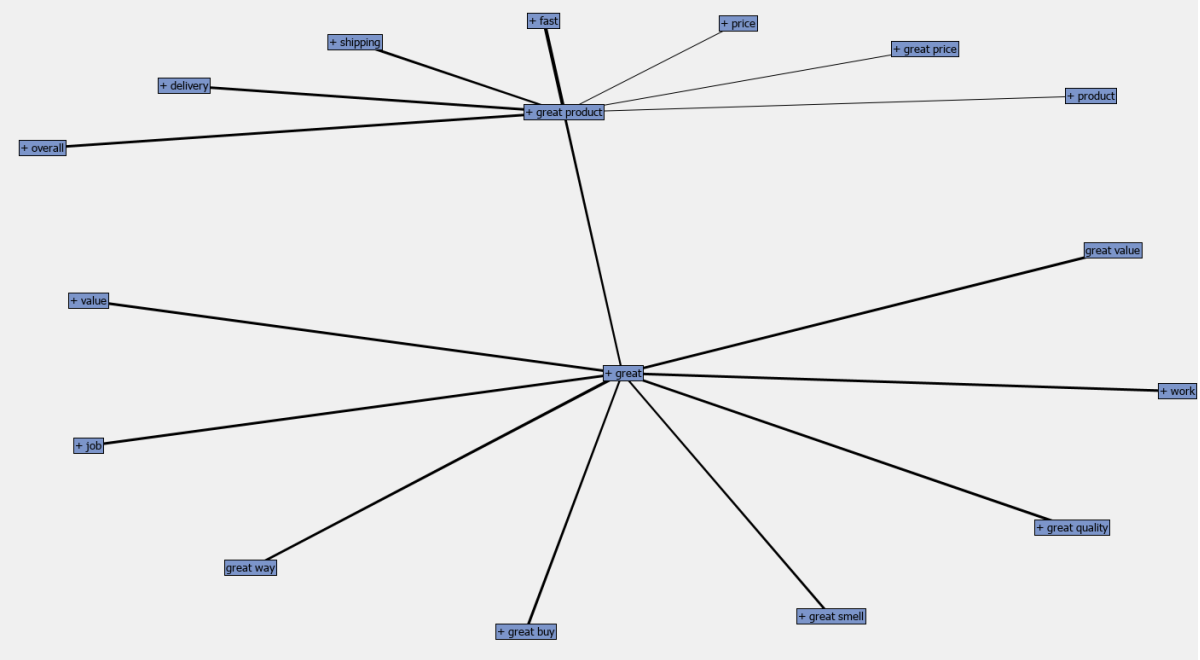


Figure 5. Concept Link for "great" word

For this study, we would like to know why people like this brand and which factors lead to success at P&G. The “great” concept link indicates that it has high association with “great product”, “great value”, “great smell”. The deeper connection from “great product” is fast, shipping, delivery. From this result, we can assume that consumers are satisfied with the quality, price and time of delivery.

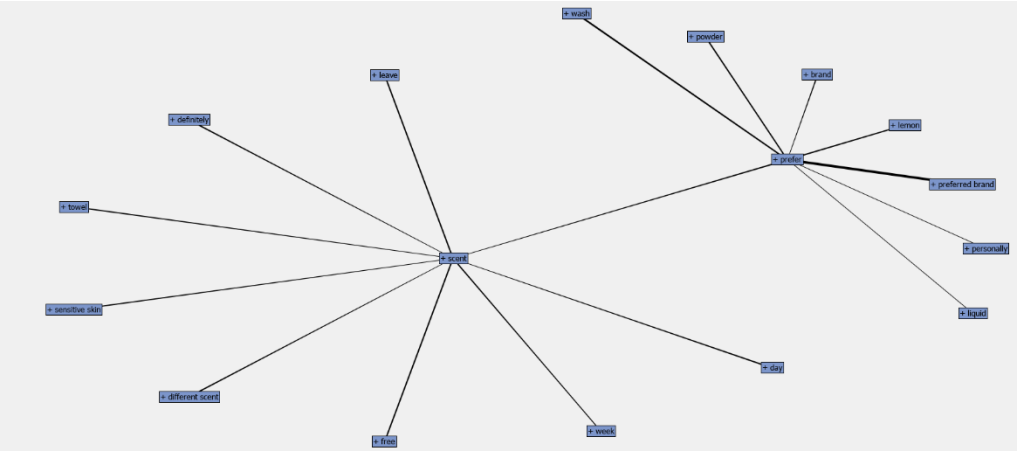


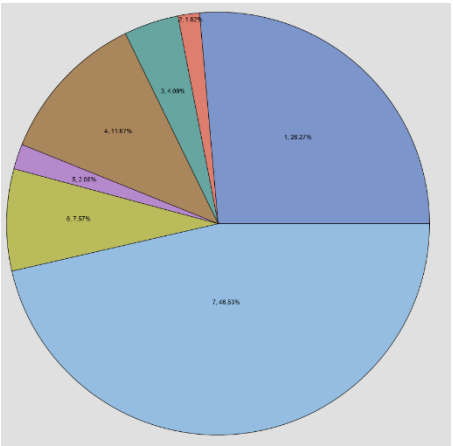
Figure 6. Concept Link for “scent” word

One of the most concern relating with chemical products in Home Care and Laundry is the scent or fragrance. Concept link in figure 6 shows word “scent” is highly associated with word “free”, “sensitive skin”, “day”, “week”, “prefer” which is highly connected to the word “lemon”. As a result, people prefer lemonade scent and it seems that the scent lasts days or weeks.

TEXT CLUSTER

Text Cluster node is connected to the Text Filter node to formulate the different groups based on the relationship among terms. Using trial-and-error method, the below properties setting is customized to get the well-separated clusters in the cluster space.

- Max SVD Dimensions: 40
- Number of clusters: 20
- Cluster Algorithm: Expectation-Maximization
- Number of Descriptive Terms: 15



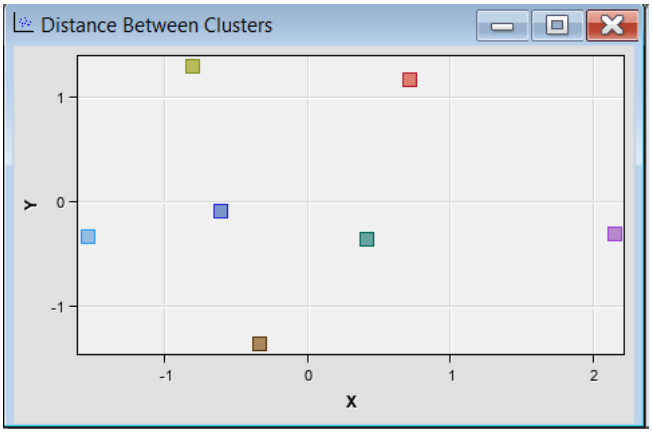


Figure 7. Text cluster result

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+purchase +bottle +little +price +easy +well +open +work +tide +mop ...	8867	26%
2	+open +strong +bottle +scent +pod +odor +little +fresh +wash +laundry ...	613	2%
3	+bad +disappoint +odor +strong +open +stick +pad +smell +purchase +dissolve ...	1380	4%
4	+buy +floor +mop +pad +easy +price +work +clean +little +bottle ...	3940	12%
5	+strong +smell +scent +odor +clothes +fresh +wash +laundry +little +purchase ...	697	2%
6	+pod +stick +tide +detergent +dissolve +laundry +easy +dishwasher +wash +clothes ...	2555	8%
7	+clean +smell +great +good +product +review +collect 'this review' +promotion +love ...	15707	47%

Figure 8. Text cluster descriptive terms output

Among seven clusters, cluster 7 and cluster 1 are very predominant. Based on descriptive terms, it is evident that there is high attention on the smell, quality, price and package feature. Moreover, cluster 7 shows that most of reviews from this cluster was collected through promotion program.

TEXT TOPIC

Text Topic node is connected to the Text Cluster node, which allows to generate topics of interest from a list of terms to gain more crucial insights.

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.155	0.012	+buy,+money,+cheap,+great,+time	36	2838
Multiple	2	0.149	0.012	+pod,+dissolve,+tide,+love,+liquid	50	2045
Multiple	3	0.136	0.012	+strong,+strong smell,+wash,+strong scent,+scent	48	1092
Multiple	4	0.146	0.013	this review,+promotion,+collect,+review,+dish	101	4367
Multiple	5	0.116	0.012	+stick,+dissolve,+plastic,+bad,+break	64	1183
Multiple	6	0.112	0.013	+box,+open,+package,+break,+leak	107	2423
Multiple	7	0.113	0.013	+floor,+pad,+mop,+wet,+money	150	2589
Multiple	8	0.117	0.013	+purchase,+dissolve,+money,+time,+wash	132	3293

Figure 9. Text Topic output

Figure 8 indicates 8 different topics which are often discussed by customers. For example, in topic 1 the term “buy”, “money”, “cheap”, “great” can be inferred that this is the good value product. The topic 6 includes some terms such as “box”, “open”, “break”, “leak”, which imply that the quality of package is not good, need to be improved. Meanwhile, topic 5 addresses some problems relating to water solubility of chemical substance such as “stick”, “dissolve”, “bad”.

SENTIMENT ANALYSIS

Statistical model

SAS® Sentiment Analysis Studio is employed to build a statistical model with a sample of 800 reviews. Among those reviews, 80% are used to train the model. As a result, the Smoothed Relative Frequency and No Feature Ranking is the best model as given below chart. The model shows that negative precision is better than positive precision with the overall precision of 85%.

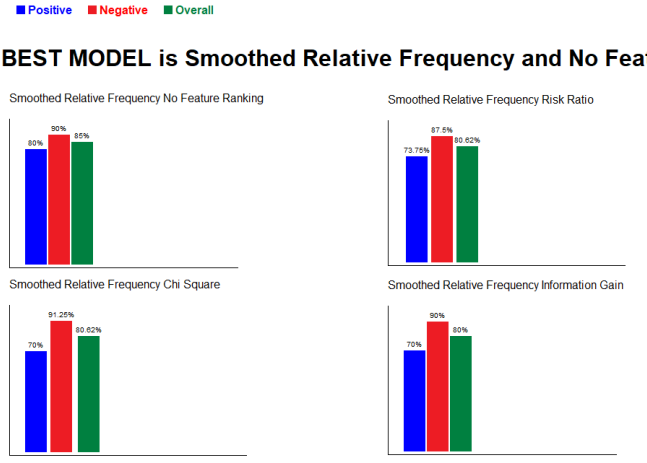


Figure 10 – Model Comparison

Test results

Test for positive reviews

From the test dataset, we read the comment and randomly picked up 400 positive and 400 negative reviews to test the model performance.

Results for selected folder:
This directory is Positive
Positive precision is 68.50%.
Number of articles:400
Number of positive articles:274
Number of negative articles:124
Number of neutral articles:2
Positive percent:68.50%.

Figure 11 – Test positive results

Among 400 reviews, there are 274 reviews predicted correctly as positive with 68.5% precision of positive comments.

Test for negative reviews

Results for selected folder:
This directory is Negative
Negative precision is 77.50%.
Number of articles:400
Number of positive articles:75
Number of negative articles:310
Number of neutral articles:15
Positive percent:18.75%.

Figure 12 – Test negative results

The model performed better in predicting negative directory with 77.5% of precision of negative reviews.

Text Rule Builder

To classify the sentiment in SAS Enterprise Miner, Text Rule Builder node is joined with Text Filter node. Text Rule Builder node enables us to examine the rule terms for positive and negative reviews. After trial-and-error, the setting with medium Generalization Error, medium Purity of Rules and high Exhaustiveness generates the best results with lowest misclassification rate.

Fit Statistics	Statistics Label	Train	Validation
ASE	Average Squared ...	0.023697	0.024618
DIV	Divisor for ASE	67518	16882
MAX	Maximum Absolut...	0.722096	0.722096
NOBS	Sum of Frequencies	33759	8441
RASE	Root Average Squ...	0.153938	0.156903
SSE	Sum of Squared E...	1599.965	415.6085
DISF	Frequency of Clas...	33759	8441
MISC	Misclassification ...	0.094108	0.10911
WRONG	Number of Wrong ...	3177	921

Figure 13 – Text Rule Builder Fit Statistics

Positive Rules

The positive rules includes some terms such as fast shipping, reliable, amaze, best price, good, brilliant, worth, clean smell. The precision of positive rules indicates the percentage of matching that term to target variable.

Some of positive rules can be shown as follows:

POSITIVE	9	fast shipping	99.59%
POSITIVE	10	reliable	99.60%
POSITIVE	11	amaze & ~hope & ~package & ~disappoint & ~first & ~cycle & ...	99.52%
POSITIVE	12	collect & ~clothe & ~dissolve & ~just ok & ~excite & ~small & ~...	99.53%
POSITIVE	13	excellent & ~fill & ~wrong & ~star & ~leak	99.48%
POSITIVE	14	best price	99.48%
POSITIVE	15	sparkly	99.49%
POSITIVE	16	great price & ~disappoint & ~dissolve	99.44%
POSITIVE	17	spotless & ~first	99.43%
POSITIVE	18	good & ~bad & ~disappoint & ~strong & ~stick & ~chemical & ~...	99.41%
POSITIVE	19	brilliant	99.41%
POSITIVE	20	worth	99.42%
POSITIVE	21	clean smell	99.42%

Figure 14 – Positive Rules

Negative Rules

The negative rules consist of some terms such as unusable, useless, dissolve but not clean and ruin, awful but not great, waste and money. The precision of negative rules shows the percentage of matching that term to target variable.

Some of negative rules are given as follows**:**

Target Value ▲	Rule #	Rule	Precision
NEGATIVE	81	dissolve & ~clean & ruin	100.0%
NEGATIVE	82	plastic & melt	100.0%
NEGATIVE	83	unusable	98.65%
NEGATIVE	84	awful & ~great	92.31%
NEGATIVE	85	useless	91.00%
NEGATIVE	86	return & ~promotion & ~always & ~clean	89.46%
NEGATIVE	87	horrible & ~clean	88.54%
NEGATIVE	88	pod & ~clean & ~love & ~easy & ~great & ~good & ~measure ...	88.79%
NEGATIVE	89	waste & money & ~dishwasher	86.92%
NEGATIVE	90	terrible & ~great & ~clean	86.68%
NEGATIVE	91	aquatic	86.82%

Figure 15 – Negative Rules

Model comparison

Predicting reviews whether positive or negative is considered as classification problem. Others predictive models such as Logistic Regression, Decision Tree, Neural Network were built and then compared to Text Rule Builder model to find the most optimal model.

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Selection Criterion: Valid: Average Squared Error	Valid: Misclassification Rate
Y	TextRule	TextRule	Text Rule Builder	Sentiment	0.024618	0.10911
	Neural	Neural	Neural Network	Sentiment	0.088244	0.114915
	Reg	Reg	Regression	Sentiment	0.09099	0.118706
	Tree	Tree	Decision Tree	Sentiment	0.093053	0.123327

Figure 16 – Model comparison

The valid misclassification rate of Text Rule Builder model is lowest among other models. Therefore, Text Rule Builder is selected as best performance model.

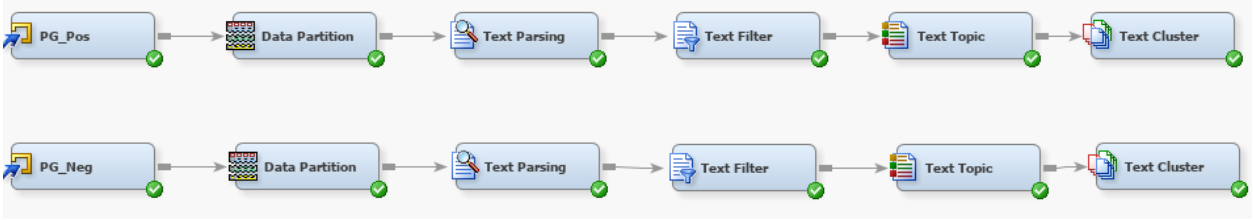
Score node

The dataset used for scoring is same as Test dataset in Sentiment Analysis Studio with 800 observations. The model shows that 377 reviews classified as positive and 131 observations are classified as negative. The Text Rule Builder in SAS Enterprise Miner predicted positive reviews better than SAS Sentiment Analysis Studio.

ANALYSIS 2 – POSITIVE AND NEGATIVE REIVEWS

In this section, all comments are divided into positive and negative dataset to analyze the effectiveness and improvement features.

1. Diagram



2. Text cluster

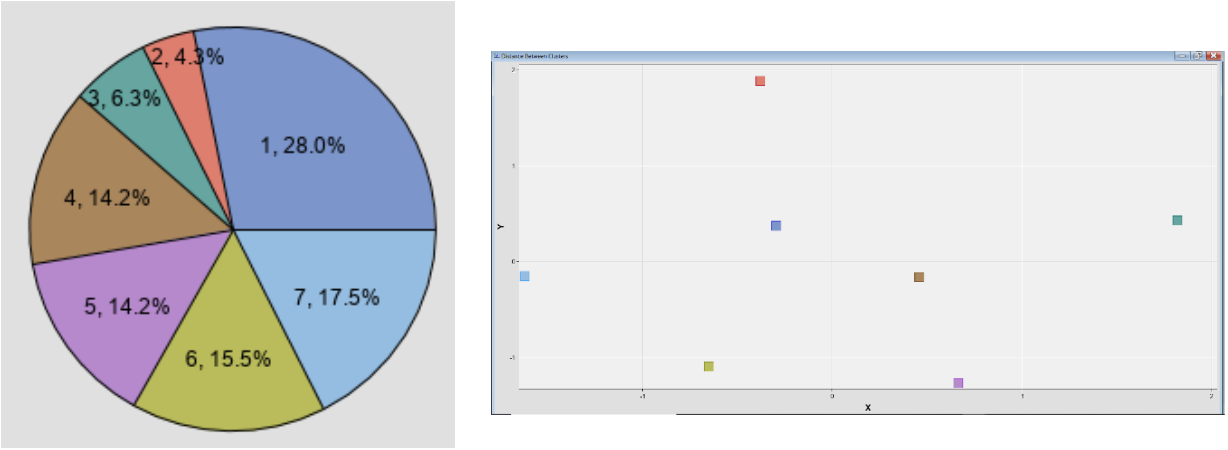


Figure 17. Text cluster pie chart and distance between clusters for positive comments

Cluster ID	Descriptive term	Percentage	Feature
1	+smell clothes +love +scent +fresh +wash +laundry +last +nice +amaze +fabric +soft +long +tide +keep	28%	smell
2	+love +scent +pod +tide swiffer +easy +refill +gain +order absolutely +stuff +thing +last +dust +store	4%	scent
3	+good +price +'good price' +'good product' +value +deal 'good value' +quality +delivery +product +stuff +order +cheap +soap +fast	6%	price
4	+work well +detergent +easy +buy +pod +tide +stain +dust +cheap swiffer +refill +store +soap +floor	14%	convinience
5	+great +product +work +price +'great product' +'great price' +value excellent +fast 'great value' +dust +delivery +order +quality shipping	14%	shipping
6	+clean +dish well +dishwasher +floor +job +water +dust swiffer +easy +soap +thing +leave cascade +stain	15%	cleaning power
7	+review +collect 'this review' +promotion +dish +dishwasher cascade +leave +soft +brand platinum +fabric +softener downy +recommend	17%	promotion campaign

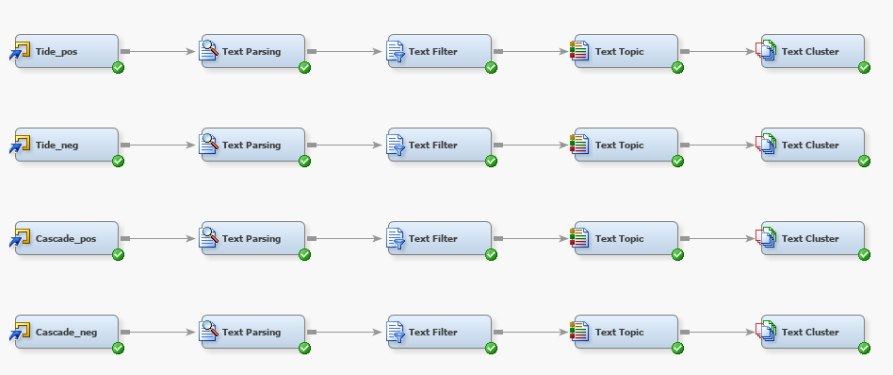
Figure 18. Cluster of positive comments

Cluster ID	Descriptive term	Percentage	Feature
1	+pod +break +dish +wet +leak +dishwasher +box +package +time +open +leave +small +floor +pad +bottle	33%	package
2	+wash clothes +review +fabric 'this review' +collect promotion +machine +softener +smell laundry +clothing +detergent +feel +nice	12%	fabric product
3	+scent clothes +pod +dissolve +tide +stain +detergent +strong +stick +load laundry +washer +plastic +clothing +leave	16%	cleaningn power
4	+smell +product +clean +good +work +buy +spray +well +strong +great +price +bad +pad +job +floor	39%	smell

Figure 19. Cluster of negative comments

ANALYSIS 3 – TIDE AND CASCADE REVIEWS

1. Diagram



2. Text cluster

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+good +price 'good product' 'good price' 'good deal' 'good value' 'good stuff' +deal +taste +product +'best detergent' value pretty quality +stuff ...	436	8%
2	+tide +pod +'tide pod' always +year +love +time +brand free +measure 'good value' +easy +container +liquid 'good deal' ...	577	11%
3	+detergent +laundry +load +'laundry detergent' +liquid +bottle convenient +little +machine +container +time +wash +scent +find +stain ...	1226	22%
4	+love +product 'this review' +collect promotion +delivery excellent +order +review +fast +scent +amaze +store +buy +price ...	818	15%
5	+clean +smell clothes +fresh +leave +nice well +wash +amaze +washer +scent +keep +machine +brand +job ...	1584	29%
6	+great +work +product +price +'great product' +'great price' 'great value' +expect value +fast well +delivery +stain +taste +deal ...	815	15%

Figure 20. Text cluster output for Tide having rating 4 & 5

Six clusters are created for Tide positive comments. Cluster 5 shows the highest frequency of percentage 29%. It shows that Tide has fresh smell and works well with machine. Customers are also interested in the load capacity and convenient package.

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+product +package +leak +box +receive +container +pod +open +bottle +'tide pod' +arrive +order +break +price +purchase ...	290	26%
2	+detergent +tide +scent +clean tide job best +love +price regular well back +good +bottle great ...	181	16%
3	+smell +strong +work +chemical +odor awful +good +house +stain well +remove +buy +bad laundry +clean ...	311	28%
4	clothes +wash +dissolve +stick clothing +water +ruin +leave +hot +machine +cycle +melt plastic completely +shirt ...	316	29%

Figure 21. Text cluster output for Tide having rating 1, 2 &3

Six clusters are formed for Tide negative comments. Cluster 1, 3 and 4 have the highest frequency percentage. It indicates that Tide has some problems with packaging which lead to leak or break during delivery. Moreover, people often complain about the strong chemical odor and the residue of detergent on clothes.

Cluster ID	Descriptive Terms	Frequency	Percentage
1	cascade +brand platinum +year +convenience best excellent +result dishwashing always +back amazon 'great deal' quality +good ...	230	3%
2	+clean +dish +pod well +rinse +wash +soap +leave +little +time +residue +stick +detergent stuff +sparkle ...	1581	23%
3	+product +great +love +'great product' +price +'good product' shipping excellent +fast +good +'great price' awesome amazon quality 'good price' ...	419	6%
4	+review 'this review' +collect promotion cascade platinum +sparkle +back +cascade +love always +year +scent +smell +result ...	3407	50%
5	+dishwasher best +detergent +'dishwasher detergent' +work +soap stuff +cheap +pod +buy amazon +liquid +hard +little +water ...	682	10%
6	+work +great +good well +price +'great price' +value +deal 'good price' 'great value' 'great deal' convenient amazon +job +store ...	509	7%

Figure 22. Text cluster output for Cascade having rating 4 & 5

Figure 22 shows six clusters for Cascade 4 & 5 rating. Cluster 4 and 2 have highest frequency. People talked much about cascade platinum product and they really love the scent and the cleaning power of Cascade, which leave little residue of detergent stuff.

Cluster ID	Descriptive Terms	Frequency	Percentage
1	+scent +strong +overpower +fresh +dryer laundry +fabric promotion 'this review' +love +collect +fragrance +chemical +review clothes ...	667	15%
2	+product +work +buy +good amazon +stop +well +price swiffer +package +great +remove +box +year +spray ...	792	18%
3	+smell +strong +nice awful +chemical +fresh +wash clothes +house +bad +great laundry +fragrance +day +overpower ...	829	19%
4	+clean +floor +pad +mop +wet swiffer +fall dirt +cloth +dry money +easy +waste +look +keep ...	474	11%
5	+pod clothes +dissolve +wash +leave +stick +detergent +load +stain +tide +water +washer +clothing +plastic +machine ...	661	15%
6	+review +bottle +box +dish +collect 'this review' promotion +dishwasher +leak +open +break +order +small +package +time ...	923	21%

Figure 23. Text cluster output for Cascade having rating 1, 2 &3

Figure 23 shows six clusters for Cascade negative rating. Cluster 6 and 3 has highest frequency. From the analysis, Cascade has the packaging problem (box, bottle, leak) and chemical smell.

CONCLUSION

The analysis leads to the following conclusions:

- Overall, customers have high satisfaction with P&G products with an average rating of 4.5. In particular, Fabric Care has an average rating of 4.47 and Home Care has average rating of 4.5.
- Fresh smell/lemonade scent, good price, convenience and fast delivery attribute to higher rating by the consumers.

- The result shows that most concerns are product packaging, chemical smell and soap residue related.
- While the Tide product gets compliments for nice smell, loading capacity and convenient package, Tide pod does not dissolve well in the water, leaving residue on clothes.
- Cascade is good at cleaning dishes but the package size and low quality of package causes a lower rating. This is also a common challenge in transportation and supply chain.
- Most reviews used in the analysis are collected by promotion program. This channel may be a good resource to capture an accurate picture of customer experience, along with online review on e-commerce platforms.

REFERENCES

Chakraborty, G., Pagolu, M., & Garla, S. (2014). Text mining and analysis: practical methods, examples, and case studies using SAS. SAS Institute

Liu, J., Sarkar, M.K., & Chakraborty, G. (2013). "Feature-Based Sentiment Analysis on Android App Reviews Using SAS® Text Miner and SAS® Sentiment Analysis Studio'. *Proceedings of the SAS Global Forum 2013*

Home & Laundry Care in US from 2020 to 2023, Statista 2019. Accessed May 2019

https://www.statista.com/outlook/60000000/109/home-laundry-care/united-states#market-arpu

COVID-19: Tracking the impact on FMCG, Retail and Media, Nielsen. Accessed March 2020

https://www.nielsen.com/us/en/insights/article/2020/covid-19-tracking-the-impact-on-fmcg-and-retail/

2019 Annual Report and Proxy Statement, pginvestor.

https://www.pginvestor.com/CustomPage/Index?KeyGenPage=1073748359

ACKNOWLEDGMENTS

We sincerely thank Dr. Goutam Chakraborty and Dr. Miriam McGaugh for their constant support and valuable direction in this research.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Trinh Phan
Oklahoma State University
Stillwater, OK 74078
Phone: (281) 771-8444
Email: trinh.phan@okstate.edu