

Capstone Group Presentation and Write up_Individual Contribution



Individual Contribution

Name of Student: Md Tajrianul Islam

Islam.mdta@northeastern.edu

ALY6980 23293 –Experiential Capstone

Summer 2021 Quarter

Submitted to –Eugenia Bastos

MPS Analytics, College of Professional Studies, Northeastern University

Capstone Group Presentation and Write up_Individual Contribution

Introduction

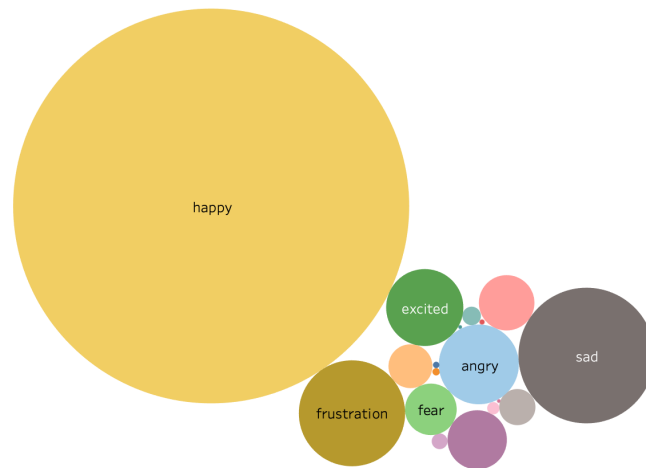
I had the chance of leading the group to complete this capstone project. Other than planning the project timeline and setting up individual deliverables, I am also a very effective person when it comes to designing research processes, doing necessary literature review and domain research. So, we all started with our EDA with the given dataset. A major role for me was to explain the sponsor deliverable and make clear ideas about our goals to my group mates. After group discussions we decided to make our target variable emotional engagement. Then my major role was to preprocess the data to make it ready for my group members to build the models on it. After the models and validations were ready I went through the codes to ensure we are on track to meet our project goals. After the codes were ready, I went ahead and prepared for the project write up and presentation along with the help of another group member.

To make sure everyone in the group is working to their strengths, we did some group exercise to ensure we as a group are on the same page about our roles and our expectations from the project. We all participated in EDA because that brought different perspectives about the existing dataset. I used Tableau to create EDA, then for data preprocessing I used Rstudio. We also used lucidspark for project kickoff meetings.

Data Analysis

We start with the EDA, then process the data, build the model and validate it. Our sponsor has provided us with a dataset that contains 15,365 rows of reviews of baby products from different pharmaceutical companies, calculated emotional engagement with PathosAI, and whether the consumer will recommend the product or not. The dataset also holds information about the date the reviews were made, what was the driver behind the sale with Ad recalls, when color and taste of the products were mentioned in the reviews, and different stages of their customer journey.

Capstone Group Presentation and Write up_Individual Contribution



The first thing we try to look at is what are the different emotions that has been registered in our data. What we see is the majority of the reviews are positive reviews, which is quite expected given the fact that they are all baby products.

happy yellow in color	happy orange color	happy carrot color	happy color even is close to	happy color is a vibrant	happy	frustration color looked gross		
						frustration cream colored		
happy artificial coloring	happy the color - much closer to breastmilk	happy green color, dark green color	happy possibly orange?	happy sane color		frustration different color	frustration yellow color	
happy different color	happy artificial color	happy make a non colored drink	happy weird coloring	happy yellow and purple		angry black color	angry green color, dark green color	disgust red coloring, red food coloring
sad green color, dark green color	sad discolored	sad green color	sad orange color	sad pale color		angry weird color	disgust pale color	
		sad green color, blue-green color	sad pink color					happy, sad
sad different color	sad off color	sad iron color	sad yellow in color					sad, angry moldy color

Capstone Group Presentation and Write up_Individual Contribution

I also wanted to see if a specific color, taste or shape have to do with the overall positivity or negativity of the product review. What we see is yellow and orange is often associated with happy feelings whereas green or dark green is associated with sadness. This resembles the color theory that we have learned in branding.

Manufacturer (group)	Recommend/Not Recommend (group)		
	Avg. Emotional Engagement		
	Null	Not Recommended	Recommended..
A2	1.98		
Abbott	1.39	0.17	1.61
All Other	1.54	0.49	2.26
Danone	1.28	1.56	1.68
Friesland Campina	-0.54		
Hain Celestial	1.25	0.20	1.48
Hero Group	2.84		
Nestle	1.54	1.39	1.74
Nutricia	1.77		
Plum Baby	1.30	-0.15	1.05
RB Health	1.43	0.21	1.58

After that we also look into how the average emotional engagement may vary a lot depending on whether a consumer recommends the product or not. One thing to notice is some of the brands have only null values for their recommendation column.

To start the analysis, I want to transform the data and make it ready to run the model. Although, for the final project I did it in a different way, where I manually encoded some specific columns for them to be used in supervised learning. To prepare the data for that, I uploaded the data, checked for spelling mistakes, null values were identified and removed, and finally data types for

Capstone Group Presentation and Write up_Individual Contribution

each of the columns were formatted. The Emotion column was divided into three categories: positive (containing one or two positive emotions), negative (containing one or two negative emotions), and unclear (containing both positive or negative). Values were classified based on that. Next, Emotional Engagement was transformed into a numerical variable.

```
library(tidyverse)
# assign each emotion a number. Negative is -5:-1 positive is 1-4
Capstone <- Capstone %>%
  mutate(Emotion_Value = if_else(Emotion == "disgust", -5,
                                if_else(Emotion %in% "angry", -4,
                                if_else(Emotion %in% "fear", -2,
                                if_else(Emotion %in% "sad", -1,
                                if_else(Emotion == "frustration", -3,
                                if_else(Emotion %in% "happy", 2,
                                if_else(Emotion %in% "excited", 4,
                                if_else(Emotion %in% "surprised", 3,0))))))))))
table(Capstone$Emotion_Value)

plot(Capstone$Date, Capstone$Emotion_Value)
sum(!is.na(Capstone$`recommend/not recommend`))
#1278 non-na's

table(Capstone$`recommend/not recommend`)

#create field for positive and neg. recommendations
Capstone$NotRecommended <- str_detect(Capstone$`recommend/not recommend`,
                                       "avoid|cannot|not|never|wouldnt")

table(Capstone$NotRecommended)
Capstone$Recommended <- str_detect(Capstone$`recommend/not recommend`,
                                   "recommend|appreciated|positive|preferred")

table(Capstone$Recommended)
```

But on the other hand, for my own research proposal I wanted to have a different target variable. which is whether the reviewer will ultimately recommend the product or not. For that also we have to follow the similar procedure of getting rid of the null values.

Capstone Group Presentation and Write up_Individual Contribution

Data Pre Processing

Id	Recommend/Not Re...	er	Emotional ..	Positive Ter...	Positive Ter...	Positive Ter...	Positive Ter...	Positive Ter...	Positive Ter...	Positive Ter...	Positive Ter...	Positive Ter...	Negative Te...
ome	1.4813278...	right	again	similar	happy	baby	so good	same				scream	f
ome	1.6728988...	able to dige...	product	loved it	gas		recommend...	satisfied	highly reco...	do well		switched	g
ome	0.5554999...	would reco...	very satisfi...	product								Null	h
enienc...	5.1871551...	very conven...	well	extra	exclusively	coupons	of course	good formu...	sometimes	like that	cheap	not cheap	n
tivity	2.0945861...	less	5 stars	likes it	even	current	tasting	recommend	more	buy it	organic	less	p
_being	1.8860739...	using	helpful	tummy	recommend...	increase	trusted	the only for...	get back	highly reco...	gentle	acid reflux	s
ome, w...	2.0047719...	love it	more	only formula	switched to...	product	gas	recommend...	highly reco...	amazing	gas		c
ome	1.0122328...	sleep	product	really helped	fussy	started usi...	would defin...	gassy	definitely r...			was very ga...	f
ome	0.9368849...	best formula	fussy	definitely w...	continue	would reco...	gassy	helps				fussy	g
ress, va...	-4.1086384...	worth the ...	same	extra	less							less	n
_being	-0.6167931...	using	again	similar	get back	good	worth the ...	fussiness				not worth	t
	1.4123636...	tried	love	fussy	definitely r...	extra	great	ready to use	easy on sto...			fussy	d
	1.4123636...	tried	love	fussy	definitely r...	extra	great	ready to use	easy on sto...			fussy	d
ome	1.0593028...	would reco...	safest	best formula								stop	f
_being	2.3168372...	tried	great	little	gassy	all	definitely w...	best thing	happy	extra	giving	smaller	v
_being	2.3168372...	tried	great	little	gassy	all	definitely w...	best thing	happy	extra	giving	smaller	v
ome, w...	1.8825990...	tried	fussiness	recommend...	spitting up	crying	gas	also	strongly re...	available	better than	never	s
ome, w...	-1.6483138...	agreed with	worth it									never	n
ome	1.21631634	recommend...	great produ...	gas	did well	drank it						uncomforta...	n
_being	1.9785175	little	recommend...	so glad	spitting up	new	gas	life saver	sensitive			spitting	li
ome	1.8755833...	coupons	as good as	spitting up	do well	worth it	use it	keep	fairly	recommend		spitting	s
_being	1.9080427...	best availa...	happy	more	even	looking for	enjoy it	product	spitting up	happily	recommend	problems	s
_being	1.9726224...	extremely	better	sleep	product	fussy	helped	gentle	good	tremendou...	gassy	extremely g...	u
ome	1.2418611...	would highl...	great									Null	h
_being	1.4152599...	ease	product	absolutely l...	finally	would defin...	gentle	miracle wor...	not gassy	definitely r...		so gassy	v
ome	1.0300417...	awesome	would reco...	everything								Null	h
ome, w...	2.1182547...	recommend...	gas	also	huge plus	would defin...	helps with	definitely r...				switched	g
_being	1.9751831...	even	love this pr...	loves	gassy	exclusively	coupons	honestly	fussy	decided to t...	to eat	hesitant	s
ome	1.1988933...	amazing for...	no problem	happy	great	worth ever...	highly reco...					switched	b
ome, w...	-1.4748533...	enough	product	instantly	recommend...	finally	well	good formu...				wouldnt rec...	h
ome, w...	1.6859848...	ease	sensitive	using	all	recommend...	gentle	less spit up	prefer over	experience	rash	less	r
ome, w...	2.16771549	tried	amazing	even	all	would reco...	works great	did great	constipation	great on it	sensitive	constipation	p

Conclusion

For the capstone project, we have applied multiple linear regression and support vector machines to predict the emotional engagement. An emotionally engaged customer is three times more likely to recommend, re-purchase and they are also less likely to be price sensitive. Also as a group we concluded that SVM is a better model to predict emotional engagement. As a group leader and a team member I think I have made a fair amount of contribution. I definitely think our project timeline could have been more organized. Also, we have a couple of member leaving the course early and not being able to join due to some personal issue, that may have slowed down our progress a little. Also, I could have been more involved in making models, but as group we decided to work on our strengths.