Individual Project Proposal

Name of the Student: Md Tajrianul Islam

Email-id: islam.mdta@northeastern.edu

ALY6980 23293 –Experiential Capstone

Summer 2021 Quarter

Submitted to –Eugenia Bastos

MPS Analytics, College of Professional Studies, Northeastern University

Individual Project Proposal

Background

Our project goal is to associate emotional reactions to increase product sales using the data provided by the sponsor of our course. One of the products offered by our sponsor Eiryam Inc is PathosPredict. It is based on PathosAI, which refers to emotional engagement, a critical factor to understanding motivations. It is a better predictor of human behavior than measurement of satisfaction and brand equity. According to the work of Coppola (2021), in 2019 nearly 1.92 billion users have bought products and/or services online. In 2020, despite the pandemic's negative impacts on people's purchasing patterns, e-commerce sales volume grew by almost 28% and constituted $4.28 trillion globally (Cramer-Flood 2021). As online sale volume develops, so does the measure of buyers and orders information that organizations can gather, store, measure, investigate and use to convey more customized and explicitly designated items and administrations to more readily address the issues of existing clients, expanding both fulfillment and degrees of consistency, and furthermore draw in new ones, expanding absolute incomes. So understanding consumer behavior has become really important for businesses in today's market. Also, it is a given truth that we human beings are also really bad at expressing ourselves. Because research shows, when asked to write a review, surveyed of why someone made a purchase the answers provided by the participant usually vary from the real reason behind those purchases. So the traditional way of using NLP can be useful for classifying simple emotions from textual data, but they aren't enough to get actionable business information. And that's where our research becomes interesting as we use the emotional engagement values produced with PathosAI.

Objectives

Our goal is to build a couple of models that will help us to maximize sales. As a group we decided to work on finding the factors that affect emotional engagement as the more invested the consumer is the more likely they are to make a purchase. But we also know, one of the best promotions for a product is word of mouth, in other words when a family or friend recommends something. So I think the way we can increase sales is by maximizing the number of people who recommend the product. My goal is to:

- Use the given data to apply multiple regression models and other methods to build a predictive model to find if a reviewer will recommend the product or not
- State the variables that pass the significance
- Analyze external data about brand perception and see if it has any effect or results in any biases

Scope

Understanding consumer behavior has become really important for businesses in today's market. A few things to note before we start our analysis:

- How is the emotional engagement related to a positive/ negative review
- Does colors, taste or shape of the product has any biases on the consumer perception
- Different business factors like Ad recall or what drove the sale may have effects on how invested they where while writing the review
- Sources for incorporating the external data, adding company perception will be interesting to see

Then, we will apply SVM, compare their accuracy and finally try to incorporate third party data to make the prediction more robust. Our project will be divided between the below phases:

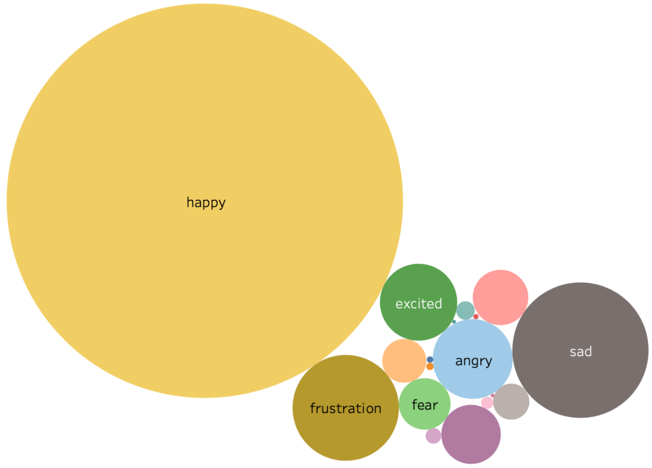
1. Literature Review
2. EDA
3. Data Preparations/ Preprocessing
4. Comparing multiple ML models
5. Hyper parameter tuning
6. Incorporating external data

Literature Review

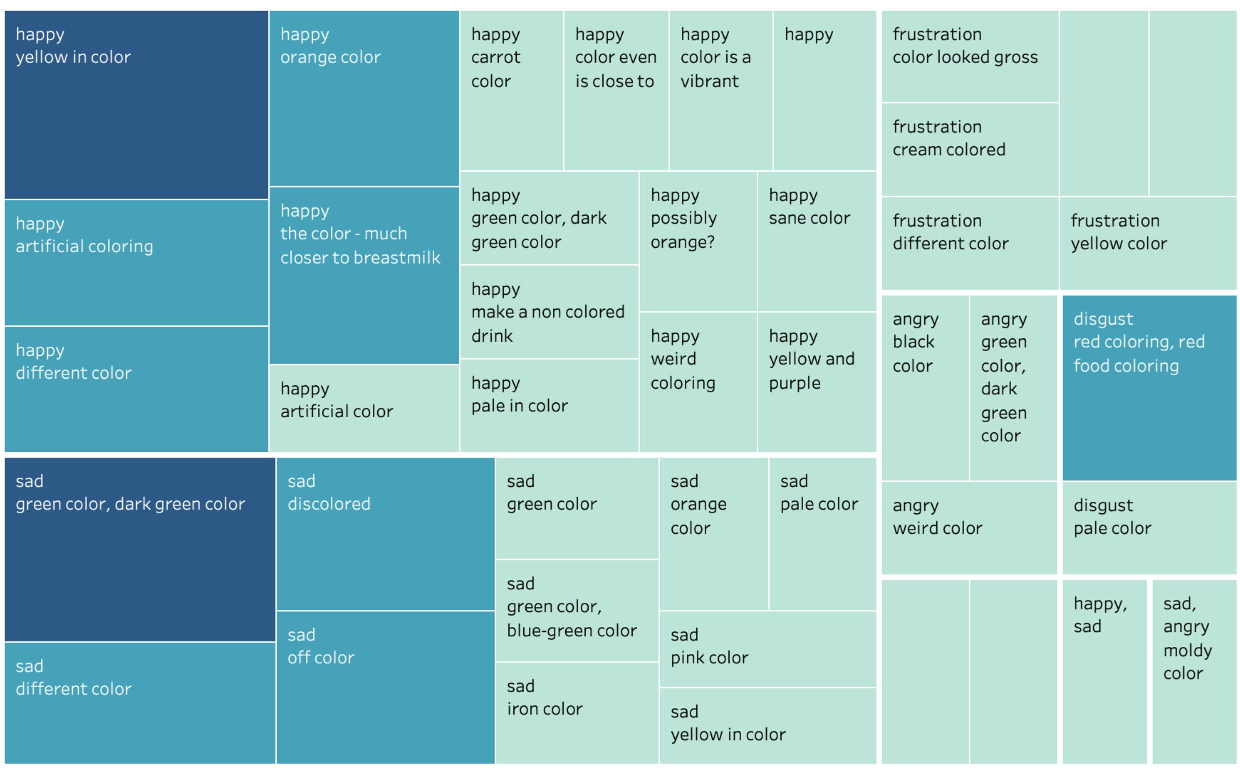
I reviewed 5 different peer reviewed articles that covered areas of how to build up a research hypothesis for consumer behavior analysis to what sort of algorithms can be used for such purposes. I at first started with reviewing smart retail technologies, which suggests three different factors that influence customer engagement behavior- performance expectancy, effort expectancy, influences (social & behavioural intention on actual behaviour). Social influence is a big factor as well which makes them trust the systems to get done with their job. The article also considers the influence of the behavioural intention on actual behaviour. For the best results the authors have included demographic variables, such as age group, and gender as control variables to allow for better delineation of the relationship proposed in our model and to provide a more rigorous test of the theoretical linkages. Then I also looked at how we can make predictive models to analyze driving behavior. As it deals with similar textual data, we learned that of all the chosen research that took on NN for DB investigation, about 57% utilized DNNs through 43% utilized traditional ANNs. The article additionally proposes mainstream execution assessments for ML models. I also looked at some purchase behavior modeling done for ecommerce websites for our research references. In this paper, they study brand purchase prediction by investigating practices, which may prompt brand buys. They make three observations. (1) they investigated genuine web based business information from numerous points. Zeroing in on clients' image buys, they figure out practices' development with time and practices' cooperation. (2) For various practices, they remove diverse time-developing elements that can fill in as pointers of clients' image buy. (3) They utilize a strategic relapse based model by changing the boundaries of time developing elements and others in two unique situations (the advancement purchase forecast and the everyday purchase number expectation) to build two tests. The investigation results show that the model utilizing three sorts of highlights plays out the best in the two situations, and the time-developing element assumes the main part among them.

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.



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.



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.

	Recommend/Not Recommend (group)		
	Avg. Emotional Engagement		
Manufacturer (group)	Null	Not Recommended	Recommend..
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 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.

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. What we found out from our group research is that the following variables are the most significant variables: satisfaction, switch, promotion, delivery, return, refund, smell, and recommendation. We also see the parameter for refund and return are the highest, meaning the quality of refund and return will ensure whether the consumer is emotionally vested in the product or not. We also observed that when emotions are less than 1 (negative) , the engagement mean is around 7.4, but when larger than 1 then it depends on whether customers felt "value for their money". If they did, the average engagement value was 9.4, representing 7% of the actual data. We did look for the external data, but it was difficult to find exact data that we can incorporate with our given dataset. We also as group concluded that SVM is a better model to predict emotional engagement. I did a little research on consumer perception about the manufacturers present in our dataset. But what I also noticed, there is no brand bias in our dataset. Because the emotional engagement has nothing to do with a specific manufacturer. I wish we had some more time to get into the problem. As we took the first 3 weeks for literature reviews and domain research, because we are doing this course in the summer semester we only had an additional 4 weeks to complete our research. If I had more time, I would personally like to invest more time in building a model that can help us predict if a reviewer will ultimately recommend the product or not.

References:

1) Coppola, D. 2021. E-commerce worldwide - Statistics and Facts. Statista. Retrieved from: https://www.statista.com/topics/871/online-shopping/#:~:text=As%20internet%20access%2 0and%20adoption,3.5%20trillion%20U.S.%20dollars%20worldwide.

2) Roy, S. K., Shabnam, S., & Singh, G. (2021). Modelling Customer Engagement Behaviour in Smart Retailing. Retrieved August 07, 2021, from https://journal.acs.org.au/index.php/ajis/article/view/2967/1057 Australasian Journal of Information Systems 2021, Vol 25, Research on User Involvement

3) Elassad, Z., Mousannif, H., Moatassime, H., & Karkouch, A. (2019, October 25). The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Retrieved August 08, 2021, from https://www.sciencedirect.com/science/article/pii/S0952197619302672

4) Dong, Y., & Jiang, W. (2018, January 23). Brand purchase prediction based on time-evolving user behaviors in e-commerce. Retrieved August 08, 2021, from https://onlinelibrary-wiley-com.ezproxy.neu.edu/doi/pdfdirect/10.1002/cpe.4882 DOI: 10.1002/cpe.4882

5) Statistical and sentiment analysis of consumer product reviews. (2017). Retrieved August 09, 2021, from https://ieeexplore.ieee.org/abstract/document/8203960