Emotional Engagement Prediction Analysis

ALY6980
Sponsored by: Eriyam Inc
Capstone Group Presentation
Aug 28, 2021
Professor: Eugenia Bastos
Group 2
Abraham Karmel, Md Tajrianul Islam, Nishay Jahangir, Sandeep Karapa Srivatsav, Yuwei Dong

EMOTIONALLY ENGAGED CUSTOMERS

- 3 times more likely to recommend
- 3 times more likely to re-purchase
- Less price sensitive
- Less likely to shop around

Problem Statement

PathosAI identifies emotional engagement (EE) as a critical factor to understanding motivations. It is a better predictor of human behavior than measurements of satisfaction and brand equity.

Goal: Associate emotional reactions to increase product sales

Research Hypothesis: Being able to understand the factors that influence EE will enable actionable information for businesses to act upon to maximize their sales

We want to propose and build a model that can tell us which attributes affect the EE value the most

Literature Review

- 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
- We human beings are really bad at expressing ourselves. Research shows, when asked to write a review or answer a survey to find why someone made a purchase, the answers provided by the participants usually vary from the real reason behind those purchases.
- Three different factors that influence customer engagement behavior- performance expectancy, effort expectancy, influences (social & behavioural intention on actual behaviour)
- As we are utilizing supervised learning the data preprocessing will include a hand-coded small set of documents for whatever variable(s) we are concerned about, depending on the EDA

Data

The data hold information about reviews of different healthcare products

- It covers important factors about consumer engagement:
    - Whether the review is positive or negative, how emotionally engage were the consumer while writing the review
    - What drove the emotion, whether the customer is satisfied, do they want to switch to a different brand
    - What sort of exposure to the brand they had, and different touchpoints of their customer journey
    - Will they recommend the product or not
- The dataset has a total of 27 attributes & 15,365 rows

```r
#import dataset
dataset1 <- read.csv(file.choose())
```

dataset1 × | Untitled1* × | dataset3 × | dataset2 × | 6020.R* ×

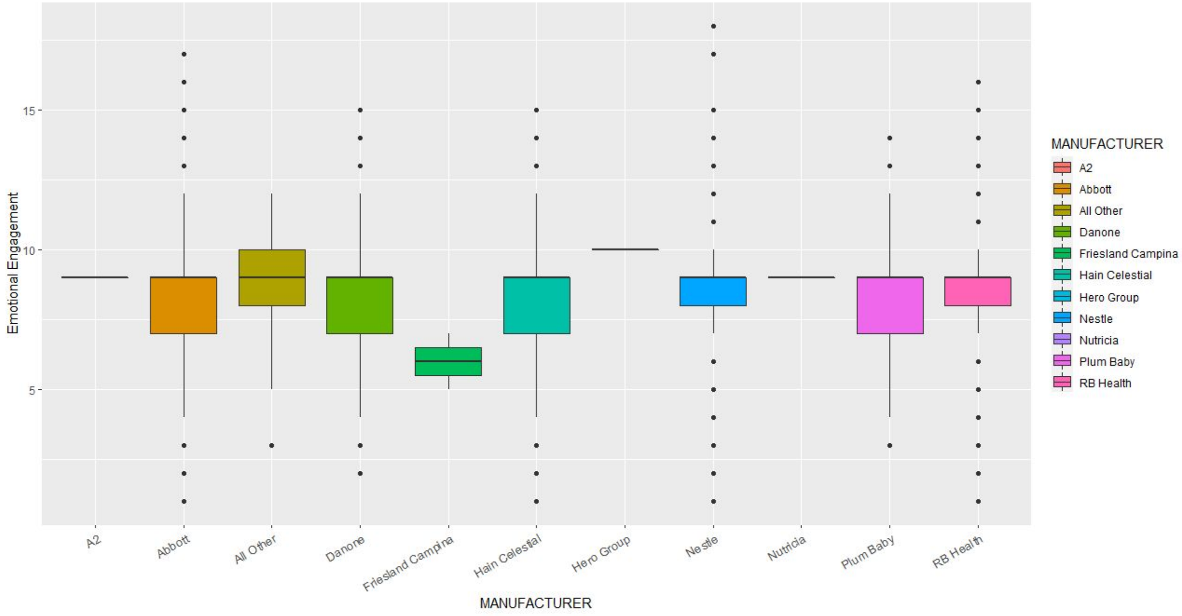
Filter

	Id	review_id	MANUFACTURER	Product	Date	Emotion	Driver	Emot Enga
1	Id	review_id	MANUFACTURER	Product	Date	Emotion	Driver	Emot
2								
3	0	139703062_327388827	RB Health	Enfamil Enspire Gentlease	2016-01-02 00:00:00	happy	well_being	1.946
4	1	7ad72964-96e6-496e-b544-7903bcd57b65_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-08 00:00:00	happy	outcome	1.481
5	2	d5ac519f-218a-4554-b666-9de96bca5af2_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-08 00:00:00	frustration	well_being	-1.04
6	3	efe60a77-b737-49aa-bd7b-92cbd3bd7a6a_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-08 00:00:00	frustration	outcome	-1.29
7	4	139703127_327388827	RB Health	Enfamil Enspire Gentlease	2016-01-09 00:00:00	happy	outcome	1.683
8	5	139703748_327388827	RB Health	Enfamil Enspire Gentlease	2016-01-12 00:00:00	happy	outcome	1.299
9	6	a2915084-9276-4c6b-9d79-310dd977bf36_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-14 00:00:00	happy	outcome	1.721
10	7	6a01f64b-52aa-458d-858b-641e3d671f7b_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-17 00:00:00	happy	outcome	1.338
11	8	16583b89-5653-4946-b5f9-994ba30c528e_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-19 00:00:00	happy	outcome	1.319
12	9	d62b3a0f-f907-426a-b737-79d0624ed404_53926215	RB Health	Enfamil Enspire Gentlease	2016-01-19 00:00:00	happy	care, well_being	1.865

```
> summary(dataset2)
      Id               review_id          MANUFACTURER          Product               Date               Emotion         
 Length:15362       Length:15362       Length:15362       Length:15362       Length:15362       Length:15362      
 Class :character   Class :character   Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character  
    Driver          Emotional Engagement Positive Terms     Negative Terms     dimensions_detected other_buy_not_buy 
 Length:15362       Length:15362         Length:15362       Length:15362       Length:15362        Length:15362      
 Class :character   Class :character     Class :character   Class :character   Class :character    Class :character  
 Mode  :character   Mode  :character     Mode  :character   Mode  :character   Mode  :character    Mode  :character  
 other_satisfied_not_satisfied other_switch_not_switch other_consider_not_consider other_advertisement_recall
 Length:15362                  Length:15362            Length:15362                Length:15362              
 Class :character              Class :character        Class :character            Class :character          
 Mode  :character              Mode  :character        Mode  :character            Mode  :character          
  promotion         customer_journey_customer_service customer_journey_delivery customer_journey_refund
 Length:15362       Length:15362                      Length:15362              Length:15362           
 Class :character   Class :character                  Class :character          Class :character       
 Mode  :character   Mode  :character                  Mode  :character          Mode  :character       
 customer_journey_return customer_journey_design_of_website    taste              color               smell          
 Length:15362            Length:15362                       Length:15362       Length:15362       Length:15362      
 Class :character        Class :character                   Class :character   Class :character   Class :character  
 Mode  :character        Mode  :character                   Mode  :character   Mode  :character   Mode  :character  
    shape           recommend/not recommend
 Length:15362       Length:15362           
 Class :character   Class :character       
 Mode  :character   Mode  :character       
```

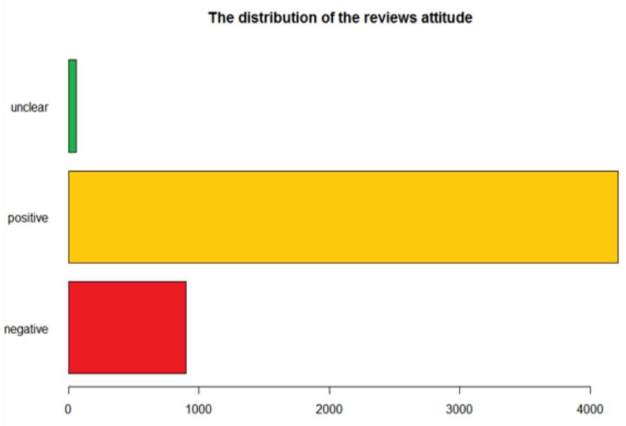
EDA

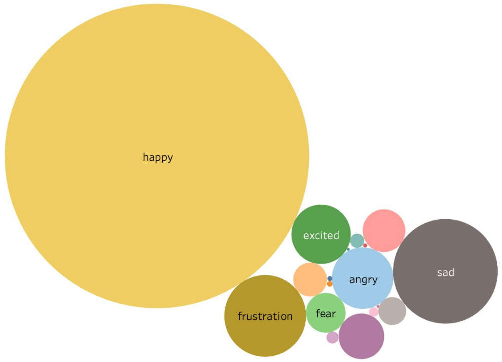
- We don’t see any brand bias in this data
- Freshland Campina is the only manufacturer that seems to have a negative EE with their consumers



EDA

- We have more positive reviews compared to negative
- This should be pretty normal for healthcare products, as they have to be FDA approved





EDA

- The attributes most correlated with EE are: satisfaction, switch, promotion, delivery, return, refund, smell, and recommendation

- When consumers don’t recommend a product, they tend to be less engaged with the review unless they had a bad experience

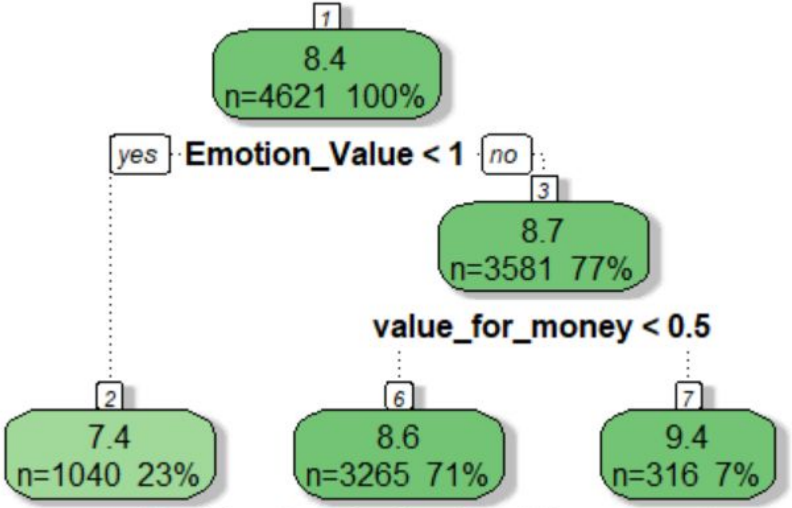
	Recommend/Not Recommend (group)		
	Avg. Emotional Engagement		
Manufacturer (group)	Null	Not Recommended	Recommend..
A2	1.98		
Abbott	1.39	0.17	1.61
All Other	1.54	0.49	2.26
Danone	1.28	1.56	1.68
Friesland Campina	-0.54		
Hain Celestial	1.25	0.20	1.48
Hero Group	2.84		
Nestle	1.54	1.39	1.74
Nutricia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

EDA

This is a decision tree for predicting Engagement, where "emotion_value" represents the numerical value of each emotion.

Negative emotions range from -1 to -5 and positive emotions are assigned from 1 to 5.

We observe 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.



Decision Tree for Emotional Engagement

Data Preprocessing

- Incorrect column names were corrected and irrelevant columns were dropped
- Null values were also identified and removed
- The data type for each column was formatted. For example, the Date column which was characteristic was converted into date type
- 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).
- Next, Emotional Engagement was transformed into a numerical variable
- For the last few columns, the null values were replaced with 0 to indicate that the review didn't mention it. If the review did mention it, a value of 1 was assigned

```
dataset2[,"Emotion"][which(dataset2[,"Emotion"]=="angry"|
                           dataset2[,"Emotion"]=="angry, frustration"|
                           dataset2[,"Emotion"]=="disgust"|
                           dataset2[,"Emotion"]=="fear"|
                           dataset2[,"Emotion"]=="frustration"|
                           dataset2[,"Emotion"]=="sad"|
                           dataset2[,"Emotion"]=="sad, angry"|
                           dataset2[,"Emotion"]=="sad, fear"|
                           dataset2[,"Emotion"]=="sad, frustration")] <- "negative"
dataset2[,"Emotion"][which(dataset2[,"Emotion"]=="excited"|
                           dataset2[,"Emotion"]=="happy"|
                           dataset2[,"Emotion"]=="happy, excited"|
                           dataset2[,"Emotion"]=="surprised"|
                           dataset2[,"Emotion"]=="surprised, happy")] <- "positive"
dataset2[,"Emotion"][which(dataset2[,"Emotion"]=="happy, angry"|
                           dataset2[,"Emotion"]=="happy, frustration"|
                           dataset2[,"Emotion"]=="happy, sad")] <- "unclear"
dataset2 <- dataset2[,-which(names(dataset2)%in%c("Driver","Positive Terms","Negative Terms","dimensions_detected"))]
```

```
dataset2[,"other_buy_not_buy"][which(dataset2[,"other_buy_not_buy"]==""))] <- 0
dataset2[,"other_buy_not_buy"][which(dataset2[,"other_buy_not_buy"]!=0)] <- 1
dataset2[,"other_satisfied_not_satisfied"][which(dataset2[,"other_satisfied_not_satisfied"]==""))] <- 0
dataset2[,"other_satisfied_not_satisfied"][which(dataset2[,"other_satisfied_not_satisfied"]!=0)] <- 1
dataset2[,"other_switch_not_switch"][which(dataset2[,"other_switch_not_switch"]==""))] <- 0
dataset2[,"other_switch_not_switch"][which(dataset2[,"other_switch_not_switch"]!=0)] <- 1
dataset2[,"other_consider_not_consider"][which(dataset2[,"other_consider_not_consider"]==""))] <- 0
dataset2[,"other_consider_not_consider"][which(dataset2[,"other_consider_not_consider"]!=0)] <- 1
dataset2[,"other_advertisement_recall"][which(dataset2[,"other_advertisement_recall"]==""))] <- 0
dataset2[,"other_advertisement_recall"][which(dataset2[,"other_advertisement_recall"]!=0)] <- 1
dataset2[,"promotion"][which(dataset2[,"promotion"]==""))] <- 0
dataset2[,"promotion"][which(dataset2[,"promotion"]!=0)] <- 1
dataset2[,"customer_journey_customer_service"][which(dataset2[,"customer_journey_customer_service"]==""))] <- 0
dataset2[,"customer_journey_customer_service"][which(dataset2[,"customer_journey_customer_service"]!=0)] <- 1
dataset2[,"customer_journey_delivery"][which(dataset2[,"customer_journey_delivery"]==""))] <- 0
dataset2[,"customer_journey_delivery"][which(dataset2[,"customer_journey_delivery"]!=0)] <- 1
dataset2[,"customer_journey_refund"][which(dataset2[,"customer_journey_refund"]==""))] <- 0
dataset2[,"customer_journey_refund"][which(dataset2[,"customer_journey_refund"]!=0)] <- -1
dataset2[,"customer_journey_return"][which(dataset2[,"customer_journey_return"]==""))] <- 0
dataset2[,"customer_journey_return"][which(dataset2[,"customer_journey_return"]!=0)] <- -1
dataset2[,"customer_journey_design_of_website"][which(dataset2[,"customer_journey_design_of_website"]==""))] <- 0
dataset2[,"customer_journey_design_of_website"][which(dataset2[,"customer_journey_design_of_website"]!=0)] <- 1
dataset2[,"taste"][which(dataset2[,"taste"]==""))] <- 0
dataset2[,"taste"][which(dataset2[,"taste"]!=0)] <- 1
dataset2[,"color"][which(dataset2[,"color"]==""))] <- 0
dataset2[,"color"][which(dataset2[,"color"]!=0)] <- 1
dataset2[,"smell"][which(dataset2[,"smell"]==""))] <- 0
dataset2[,"smell"][which(dataset2[,"smell"]!=0)] <- 1
dataset2[,"shape"][which(dataset2[,"shape"]==""))] <- 0
dataset2[,"shape"][which(dataset2[,"shape"]!=0)] <- 1
dataset2[,"recommend/not recommend"][which(dataset2[,"recommend/not recommend"]==""))] <- 0
dataset2[,"recommend/not recommend"][which(dataset2[,"recommend/not recommend"]!=0)] <- 1
```

Model

Convert to numeric

```
#convert to numeric
dataset3 <- dataset2
dataset3[,"Emotion"][which(dataset3[,"Emotion"]=="negative")] <- -1
dataset3[,"Emotion"][which(dataset3[,"Emotion"]=="positive")] <- 1
dataset3[,"Emotion"][which(dataset3[,"Emotion"]=="unclear")] <- 0
dataset3[,"Emotion"] <- as.numeric(dataset3[,"Emotion"])
dataset3[,"other_buy_not_buy"] <- as.numeric(dataset3[,"other_buy_not_buy"])
dataset3[,"other_satisfied_not_satisfied"] <- as.numeric(dataset3[,"other_satisfied_not_satisfied"])
dataset3[,"other_switch_not_switch"] <- as.numeric(dataset3[,"other_switch_not_switch"])
dataset3[,"other_consider_not_consider"] <- as.numeric(dataset3[,"other_consider_not_consider"])
dataset3[,"other_advertisement_recall"] <- as.numeric(dataset3[,"other_advertisement_recall"])
dataset3[,"promotion"] <- as.numeric(dataset3[,"promotion"])
dataset3[,"customer_journey_customer_service"] <- as.numeric(dataset3[,"customer_journey_customer_service"])
dataset3[,"customer_journey_delivery"] <- as.numeric(dataset3[,"customer_journey_delivery"])
dataset3[,"customer_journey_refund"] <- as.numeric(dataset3[,"customer_journey_refund"])
dataset3[,"customer_journey_return"] <- as.numeric(dataset3[,"customer_journey_return"])
dataset3[,"customer_journey_design_of_website"] <- as.numeric(dataset3[,"customer_journey_design_of_website"])
dataset3[,"taste"] <- as.numeric(dataset3[,"taste"])
dataset3[,"color"] <- as.numeric(dataset3[,"color"])
dataset3[,"smell"] <- as.numeric(dataset3[,"smell"])
dataset3[,"shape"] <- as.numeric(dataset3[,"shape"])
dataset3[,"recommend/not recommend"] <- as.numeric(dataset3[,"recommend/not recommend"])
```

Weight the value

```
dataset3$a <- dataset3[,"Emotion"]*dataset3[,"other_buy_not_buy"]
dataset3$b <- dataset3[,"Emotion"]*dataset3[,"other_satisfied_not_satisfied"]
dataset3$c <- dataset3[,"Emotion"]*dataset3[,"other_switch_not_switch"]
dataset3$d <- dataset3[,"Emotion"]*dataset3[,"other_consider_not_consider"]
dataset3$e <- dataset3[,"Emotion"]*dataset3[,"other_advertisement_recall"]
dataset3$f <- dataset3[,"Emotion"]*dataset3[,"promotion"]
dataset3$g <- dataset3[,"Emotion"]*dataset3[,"customer_journey_customer_service"]
dataset3$h <- dataset3[,"Emotion"]*dataset3[,"customer_journey_delivery"]
dataset3$i <- dataset3[,"Emotion"]*dataset3[,"customer_journey_refund"]
dataset3$j <- dataset3[,"Emotion"]*dataset3[,"customer_journey_return"]
dataset3$k <- dataset3[,"Emotion"]*dataset3[,"customer_journey_design_of_website"]
dataset3$l <- dataset3[,"Emotion"]*dataset3[,"taste"]
dataset3$m <- dataset3[,"Emotion"]*dataset3[,"color"]
dataset3$n <- dataset3[,"Emotion"]*dataset3[,"smell"]
dataset3$o <- dataset3[,"Emotion"]*dataset3[,"shape"]
dataset3$p <- dataset3[,"Emotion"]*dataset3[,"recommend/not recommend"]
```

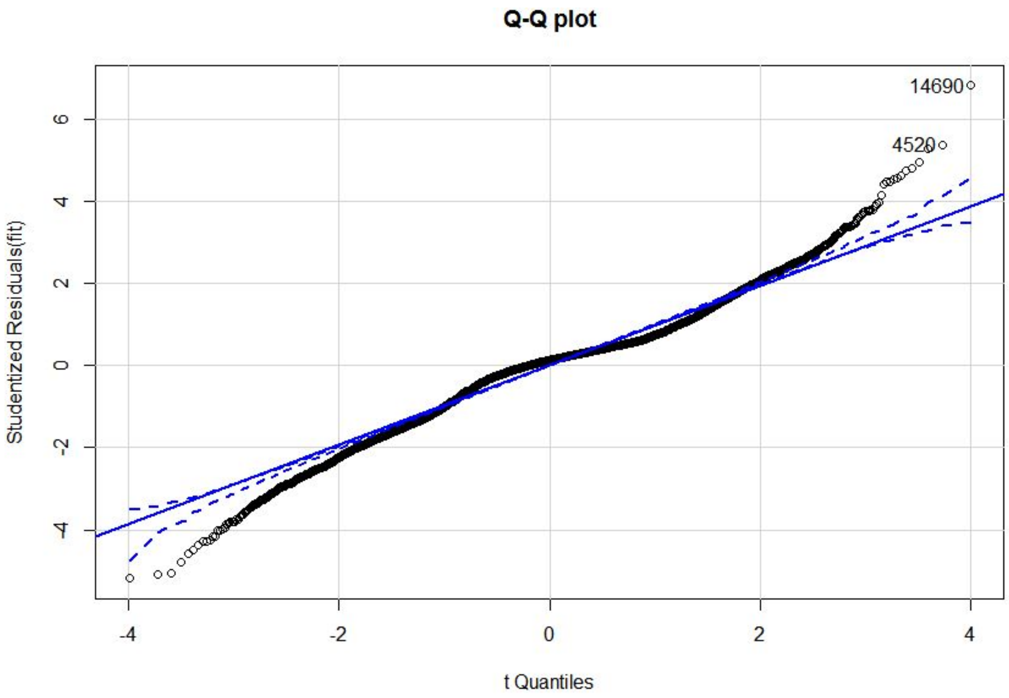
Drop irrelevent columns

```
dataset4 <- dataset3[,-which(names(dataset3)%in%c("MANUFACTURER","Product ","Date","Emotion","other_buy_not_buy"
                                                   ,"other_satisfied_not_satisfied","other_switch_not_switch"
                                                   ,"other_consider_not_consider","other_advertisement_recall"
                                                   ,"promotion","customer_journey_customer_service"
                                                   ,"customer_journey_delivery","customer_journey_refund"
                                                   ,"customer_journey_return","customer_journey_design_of_website"
                                                   ,"taste","color","smell","shape","recommend/not recommend"))]
```

Multiple Linear Regression

```
> fit<-lm(dataset4$`Emotional Engagement`~.,data=dataset4)
> summary(fit)

Call:
lm(formula = dataset4$`Emotional Engagement` ~ ., data = dataset4)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.9382 -0.7000  0.2055  0.7679 10.4347 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.361126   0.013370 101.804  < 2e-16 ***
a            0.016722   0.161296   0.104   0.9174    
b            0.278250   0.042219   6.591 4.52e-11 ***
c            0.456510   0.049019   9.313  < 2e-16 ***
d            0.298585   0.116285   2.568   0.0102 *  
e           -0.009314   0.285177  -0.033   0.9739    
f            0.431570   0.059238   7.285 3.36e-13 ***
g            0.190395   0.226374   0.841   0.4003    
h            0.473290   0.072451   6.533 6.67e-11 ***
i            0.433621   0.207740   2.087   0.0369 *  
j            0.551864   0.121626   4.537 5.74e-06 ***
k           -0.028090   0.372084  -0.075   0.9398    
l            0.324144   0.044734   7.246 4.50e-13 ***
m            0.362167   0.209168   1.731   0.0834 .  
n            0.417461   0.101969   4.094 4.26e-05 ***
o            0.228981   0.334923   0.684   0.4942    
p            0.330724   0.044667   7.404 1.39e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.532 on 15345 degrees of freedom
Multiple R-squared:  0.03415,	Adjusted R-squared:  0.03314 
F-statistic: 33.91 on 16 and 15345 DF,  p-value: < 2.2e-16
```

Hypothesis test



The data conforms to the normality hypothesis.

Hypothesis test

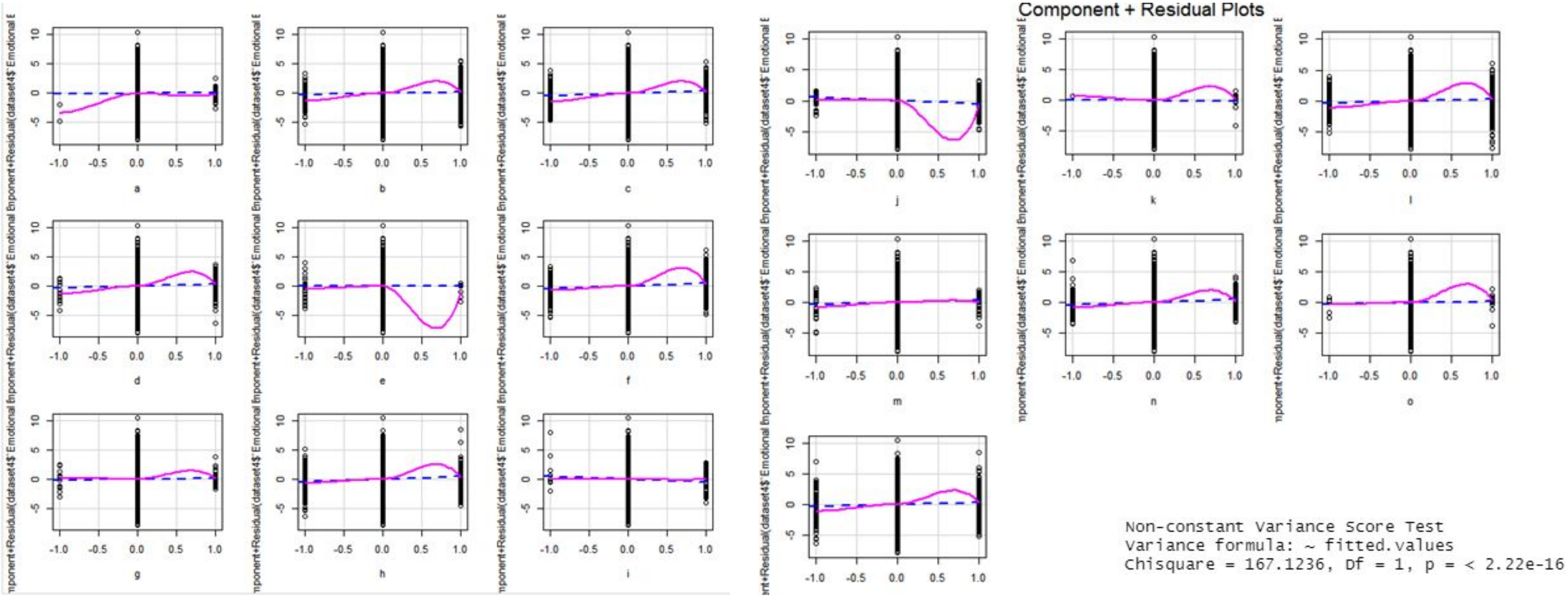
```
> durbinWatsonTest(fit)
 lag Autocorrelation D-W Statistic p-value
   1      0.01228086      1.975429   0.136
 Alternative hypothesis: rho != 0
```

D-W test.

P=0.126>0.05

There is no autocorrelation between variables and they are independent.

Hypothesis test



Residual Plots and homoscedasticity

Optimized Model

```
Call:
lm(formula = dataset4$`Emotional Engagement` ~ b + c + f + h + 
    j + i + n + p, data = dataset4)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.7801 -0.7075  0.2097  0.7692 10.4175 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.37833    0.01317 104.631  < 2e-16 ***
b            0.28211    0.04223   6.680 2.46e-11 ***
c            0.46841    0.04905   9.550  < 2e-16 ***
f            0.45070    0.05927   7.604 3.04e-14 ***
h            0.48828    0.07250   6.735 1.69e-11 ***
j            0.57158    0.12172   4.696 2.68e-06 ***
i            0.45010    0.20806   2.163   0.0305 *  
n            0.48651    0.10175   4.781 1.76e-06 ***
p            0.34971    0.04467   7.829 5.24e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.535 on 15353 degrees of freedom
Multiple R-squared:  0.03006,	Adjusted R-squared:  0.02955 
F-statistic: 59.47 on 8 and 15353 DF,  p-value: < 2.2e-16
```

All independent variables passed the significance test.

Support Vector Regression

```
> svm.model1 <- svm(dataset4$`Emotional Engagement`~.,data=dataset4,type = "eps-regression",kernel = "radial")
> summary(svm.model1)

Call:
svm(formula = dataset4$`Emotional Engagement` ~ ., data = dataset4, type = "eps-regression", kernel = "radial")


Parameters:
   SVM-Type:  eps-regression
 SVM-Kernel:  radial
       cost:  1
      gamma:  0.0625
    epsilon:  0.1


Number of Support Vectors:  13064
```

Comparison

```
> data.frame( R2 = rsquare(fit1, dataset4),
+             RMSE = rmse(fit1,dataset4),
+             MAE = mae(fit1, dataset4))
          R2     RMSE      MAE
1 0.03005528 1.534814 1.117195
> data.frame( R2 = rsquare(svm.model1, dataset4),
+             RMSE = rmse(svm.model1,dataset4),
+             MAE = mae(svm.model1, dataset4))
          R2     RMSE      MAE
1 0.04550749 1.530015 1.061343
```

SVM(SVR) has the lower RMSE

Conclusion

- SVM(SVR) has the lower RMSE, which means it is better than Multiple Linear Regression Model.
- The multi-class logistic model exhibits 72.77% to predict the consumer emotion.
- The variables that pass the significance test are b,c,f,h,j,i,n, and p, which are satisfaction, switch, promotion, delivery, return, refund, smell, and recommendation, respectively.
- The parameter values of return and refund are largest, which means the quality of returns and refunds will have a significant influence on the Emotional Engagement. We therefore suggest that enterprises carry out system reform in these aspects to meet the needs of customers to improve customer emotional engagement.
- In the remaining variables, the delivery of the goods, the smell of the products, preferential policies, and the comparison to other products are all variables with large parameters. This indicates that enterprises should focus on improving quality in these areas as well as to promote customer EE.