Assignment 7:  Sentiment Analysis

Predict 453

Section 55

Spring Quarter

School of Continuing Studies

Northwestern University

Daniel Prusinski

In compliance for Master of Science Predictive Analytics

Bachelor of Science Business Marketing

Data Analyst

US Bank

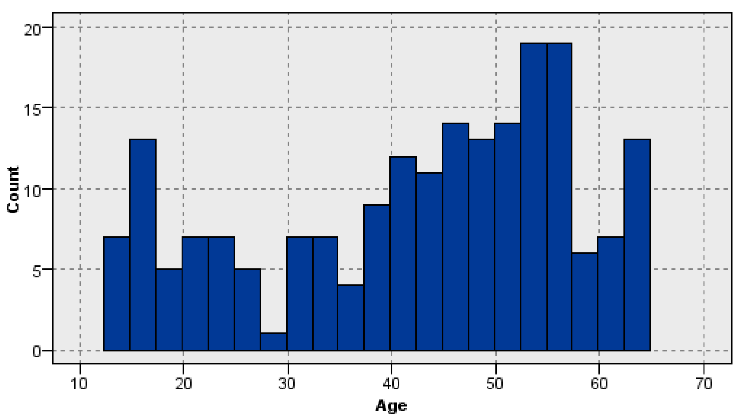
220 S 6th St

Minneapolis, MN

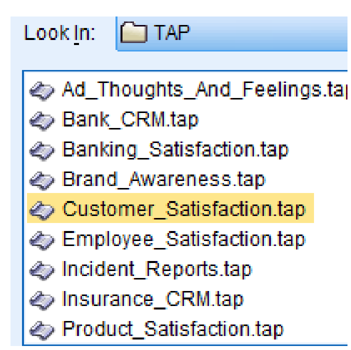
# Sentiment Analysis

While text analytics has many applications for businesses and organizations, sentiment analysis is often the pinnacle application for adding value. In data mining, numbers need to be cleaned and appropriated to corresponding vectors. Likewise, in text analytics words need to be assigned to a proper sentiment and category. This process takes strategy based on the size of the data of which one is working. Detailed throughout this exercise is the strategy and implementation of my sentiment analysis on the Car Rental Demo Data found in IBM SPSS Modeler.

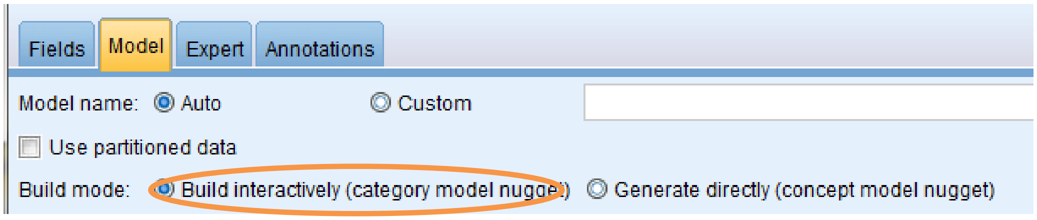
A total of 200 different individual's comments were analyzed. The text data was found within the Customer_Service vector and the histogram bellow shows the distribution of age among the individuals that submitted feedback. For this group, the average age is 42 years old with a standard deviation of 14.74. As it can be seen, Age has a negative skew. There are 115 males and 85 females. 130 individuals do not own a car and 70 do own a car.



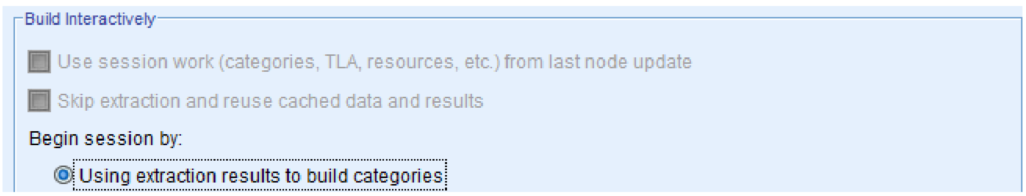

Before delving into the text analysis specifically focusing on sentiment, it was helpful to gather perspective behind the text. Given that the focus was sentiment, I utilized the Customer_Satisfaction Text Analysis Package (TAP) as my base to start from in regard to assigning sentiment.



There are different options for building a sentiment analysis, and I went with following:



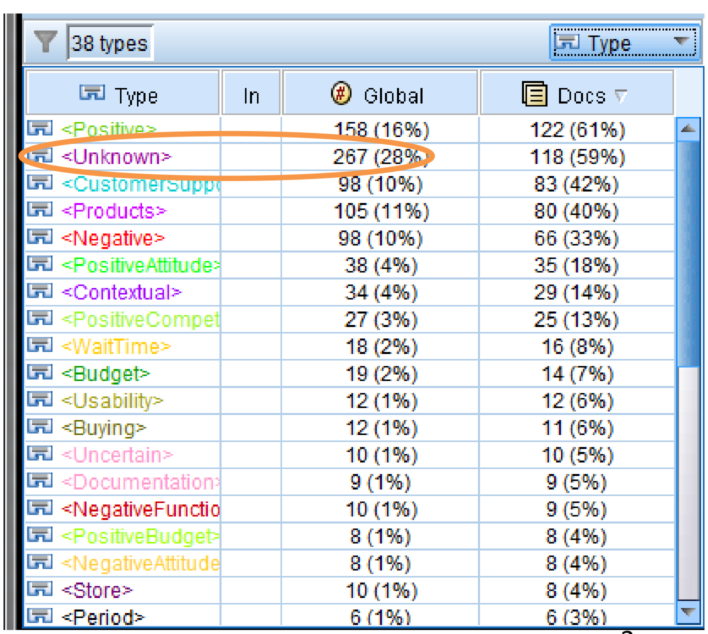

I selected the Build interactively mode in an effort to customize and tweak the analysis.




The Using extraction results to build categories was selected with the same intent of customization and interaction with the analysis.



Initially, there were 38 types and 400 concepts when the model was run. The goal of my analysis was to start small with types and work up to categories. 38 types encompassed sentiment, and is more manageable starting place than the 400 concepts. As seen below, Unknown sentiment represents about 28% of the text found in 59% of the documents. Analyzing these statements is a great starting point to customize the analysis to reflect the context of specific statements. The next step is


38 types — Type

| Type | In | Global | Docs |
|---|---|---|---|
| <Positive> | | 158 (16%) | 122 (61%) |
| <Unknown> | | 267 (28%) | 118 (59%) |
| <CustomerSuppo | | 98 (10%) | 83 (42%) |
| <Products> | | 105 (11%) | 80 (40%) |
| <Negative> | | 98 (10%) | 66 (33%) |
| <PositiveAttitude> | | 38 (4%) | 35 (18%) |
| <Contextual> | | 34 (4%) | 29 (14%) |
| <PositiveCompet | | 27 (3%) | 25 (13%) |
| <WaitTime> | | 18 (2%) | 16 (8%) |
| <Budget> | | 19 (2%) | 14 (7%) |
| <Usability> | | 12 (1%) | 12 (6%) |
| <Buying> | | 12 (1%) | 11 (6%) |
| <Uncertain> | | 10 (1%) | 10 (5%) |
| <Documentation> | | 9 (1%) | 9 (5%) |
| <NegativeFunctio | | 10 (1%) | 9 (5%) |
| <PositiveBudget> | | 8 (1%) | 8 (4%) |
| <NegativeAttitude | | 8 (1%) | 8 (4%) |
| <Store> | | 10 (1%) | 8 (4%) |
| <Period> | | 6 (1%) | 6 (3%) |

to filter Unknown and begin to work through the concepts.

| Concept | In | Global | Docs | Type |
|---|---|---|---|---|
| service | | 14 (1%) | 14 (7%) | <Unknown> |
| experience | | 11 (1%) | 11 (6%) | <Unknown> |
| company | | 10 (1%) | 10 (5%) | <Unknown> |
| process | | 6 (1%) | 6 (3%) | <Unknown> |
| rental | | 5 (1%) | 5 (3%) | <Unknown> |
| insurance | | 4 (0%) | 4 (2%) | <Unknown> |
| pick up | | 4 (0%) | 4 (2%) | <Unknown> |
| airport | | 4 (0%) | 4 (2%) | <Unknown> |

As seen above, there were concepts that occurred more than once, and specifying a sentiment direction will increase clarity. The term service occurred 14 times and only one of the 14 occurrences was negative. I marked that term negative and the rest were marked as positive. Assigning these terms to a specific type will enrich the analysis. Concepts that were linked to a specific sentiment included the following: insurance – negative, service – positive, experience – positive, process – positive. The rest of the concepts account for less than 1 percent of the analysis and moving on to a different refinement strategy is prudent. Through mapping the above concepts to categories and assigning sentiment, the overall analysis was enhanced by over 15 percent based on the global numbers of concepts that changed.

They provided good service. It was easy to make on-line reservations, the airport bus took me directly and quickly to the car, gas tank filled, customer service friendly and helped with directions and luggage.

For the most part, they it's been really good service. I usually rent from the airports because it's much cheaper. Once they charged me over and I had to hunt down the accounting department to get my money back. it was a pain.

I HATE how they pressure you to buy add-on service.

They provided the best service thus far. They are very friendly and helpful. I have rented from other companies before both cheaper and more expensive.

the service at the airport was very convenient. no lines at the service desk.

Service was fast. Car was parked close-by.

My last rental experience for leisure was fine -- no service issues to speak of.

Synonym customization is another technique that clarifies analysis. I sorted the concepts by Global occurrence, descending, and looked for opportunities to link concepts together. At the category level, SPSS does a solid job of grouping together the like terms. Earlier in the analysis, I selected the Customer_Satisfaction dictionary which has a built in library that makes this process relatively complete at the category level. There are 28 categories that all the concepts fall into based on the uploaded dictionary. Given that this analysis is focused on customer sentiment, looking through the categories and re-assigning categories to proper sentiments will aid in a clear analysis. The category "Well" was not assigned a positive sentiment but all the comments were positive.

| egory | Descriptors | Docs |
|---|---|---|
| fx [ sophisticated] | | 0 |
| fx [* options * & <Positive>] | | 0 |
| fx [ <PositiveFeeling> & !( lightweight )] | | 0 |
| fx [ creative] | | 0 |
| well | 7 | 2 |
| well-designed | | 0 |
| well-packaged | | 0 |
| well-run | | 0 |
| well-made | | 0 |
| well | | 2 |
| well-staffed | | 0 |
| well-documented | | 0 |
| Neg: Service Dissatisfaction | 42 | 0 |
| fx [ live assistance] | | 0 |
| fx [ <Buying> & !( advertisement ) & ( <Negative> \| better )] | | 0 |
| fx [ <CustomerSupport> & ( <Negative> \| better \| no \| few )] | | 0 |
| fx [ <NegativeCompetence>] | | 0 |
| fx [ <WaitTime> & !( <Positive> ) & !( <PositiveAttitude> ) & !( no )] | | 0 |

I re-inserted the 'Well' category within the "Pos: Product: Design-Features" category. As seen above, this addition better categorizes the text to reflect the sentiment. The categories "upgraded", "timely", "good", and "easy" were all assigned to positive sentiments. I then merged all the categories into one of two potential categories, positive and negative.

| egory | Descriptors | Docs |
|---|---|---|
| All Documents | - | 200 |
| Uncategorized | - | 106 |
| No concepts extracted | - | 1 |
| Other: Don't Know | 9 | 0 |
| Other: No Experience-Does Not Apply | 5 | 0 |
| Other: No Like-No Dislike | 4 | 0 |
| Overall Positive | 380 | 83 |
| Overall Negative | 72 | 12 |

Overall there are 200 surveys. 105 remain uncategorized, 83 are positive and 12 are negative. I then read through the uncategorized and assigned the surveys to one of three potential categories, positive, negative, other. After categorizing all the documents, the output is listed below.

| Category | Descriptors | Docs |
|---|---|---|
| All Documents | - | 200 |
| Uncategorized | - | 10 |
| No concepts extracted | - | 1 |
| Overall Positive | 427 | 170 |
| Overall Negative | 101 | 110 |
| Other Overall | 18 | 0 |

170 of the documents are marked as positive, and 110 are marked as negative. I will now need to work through the categories to better classify the documents. After multiple iterations, the documents have been sorted into the following categories.

| Category | Descriptors | Docs |
|---|---|---|
| All Documents | - | 200 |
| Uncategorized | - | 14 |
| No concepts extracted | - | 1 |
| Overall Positive | 103 | 141 |
| Overall Negative | 93 | 58 |
| Other Overall | 52 | 110 |

Within positive, the following terms were found:

| Category | Descriptors | Docs |
|---|---|---|
| Overall Positive | 103 | 141 |
| knowledgeable | | 6 |
| fast | | 20 |
| helpful | | 15 |
| friendly | | 21 |
| pleasant | | 2 |
| accommodating | | 1 |
| smooth | | 1 |
| detailed | | 1 |
| satisfied | | 4 |
| excellent | | 16 |
| university price | | 1 |
| answer | | 2 |
| convenient | | 2 |
| courteous | | 7 |
| appropriate | | 3 |
| Positive Descriptors | 6 | 60 |
| Pos: Product: Design-Features | 57 | 2 |

The concepts found in positive accurately represent positive sentiment from the original surveys. The overall number quantifies the positive expressions individuals wrote.

The negative sentiment is shown below:

| Overall Negative | 48 | 58 |
|---|---|---|
| staff to accommodate travelers | | 1 |
| wait time | | 6 |
| hurry | | 1 |
| turn | | 1 |
| need more | | 3 |
| concern | | 1 |
| slow | | 4 |
| internet | | 2 |
| dislike | | 8 |
| bad | | 9 |
| frustrating | | 1 |
| too long | | 13 |
| 3 years | | 1 |
| price | | 5 |
| system | | 1 |
| provider | | 1 |
| stupid | | 1 |

Within this category, all the negative sentiments are captured. There were terms like "staff to accommodate travelers" that needed to be re-classified in order to capture the correct sentiment. Other represents terms and concepts that were indifferent for sentiment analysis. For example, the concept "directions" contained comments about people asking for direction. This did not directly tie into sentiment and is classified as other. The category is seen below:

| Other Overall | 31 | 106 |
|---|---|---|
| car | | 50 |
| map | | 2 |
| staff | | 16 |
| flight | | 4 |
| line | | 9 |
| company | | 10 |
| technical support | | 8 |
| people | | 7 |
| directions | | 6 |
| location | | 4 |
| change | | 2 |
| car to the hotel | | 1 |
| greater | | 2 |
| directions to my hotel | | 1 |
| color of the car | | 1 |
| child seat | | 1 |
| size of car | | 1 |

Putting these concepts into 'other' allows one to gather a more complete analysis.

At the macro level, 200 different documents have been analyzed for sentiment. Within the 'Positive' category 141 documents conveyed some form of positive sentiment. The following terms were used the most: Overall sentiment shows that

| Term | # in Documents |
|---|---|
| Good | 53 |
| Friendly | 21 |
| Fast | 20 |
| Excellent | 16 |
| No Problem | 16 |
| Helpful | 15 |

141/200= .705 or 71% of the total surveys had positive sentiment. Based on what the company values, terms like friendly, fast, excellent, and helpful are accurate, positive adjectives of the company. This information could be used for future marketing campaigns or assessing internal candidates performances.

| Term | # in Documents |
|---|---|
| Slow/Too Long | 32 |
| Bad/Dislike | 17 |
| Pushy | 7 |
| Price | 5 |

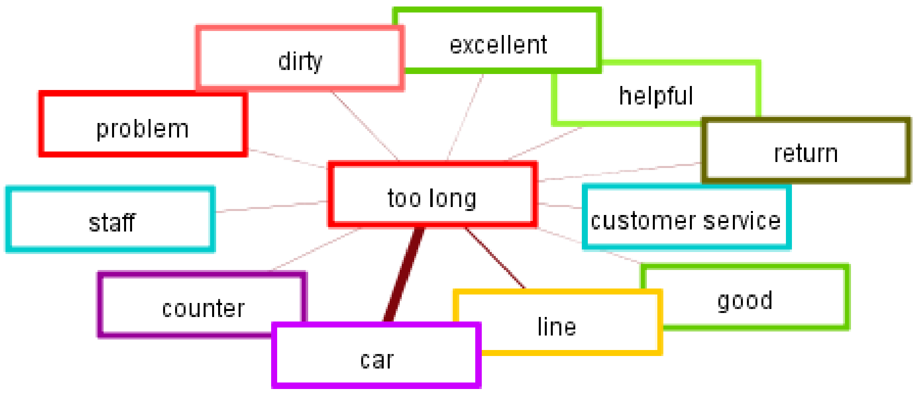
Roughly, 62/200 = .31 or 31% of the total surveys had negative sentiment. 10% of survey respondents state that the service was fast, but 16% state it takes too long. Roughly 6% more of customers think that the rental service is slow and this could be an area of improvement for the organization. The 'goods' outweighed the 'bads' by quite a bit. The concept insurance was originally a neutral term, but after further investigation it was found that customers felt agents were being pushy about selling add-ons like insurance. This is another point that would help train agents to execute in a manner that grows the business. The car rental organization has far more positive sentiment than negative, roughly double. From this macro sentiment analysis, there are key opportunities for this organization to work on to become even better.

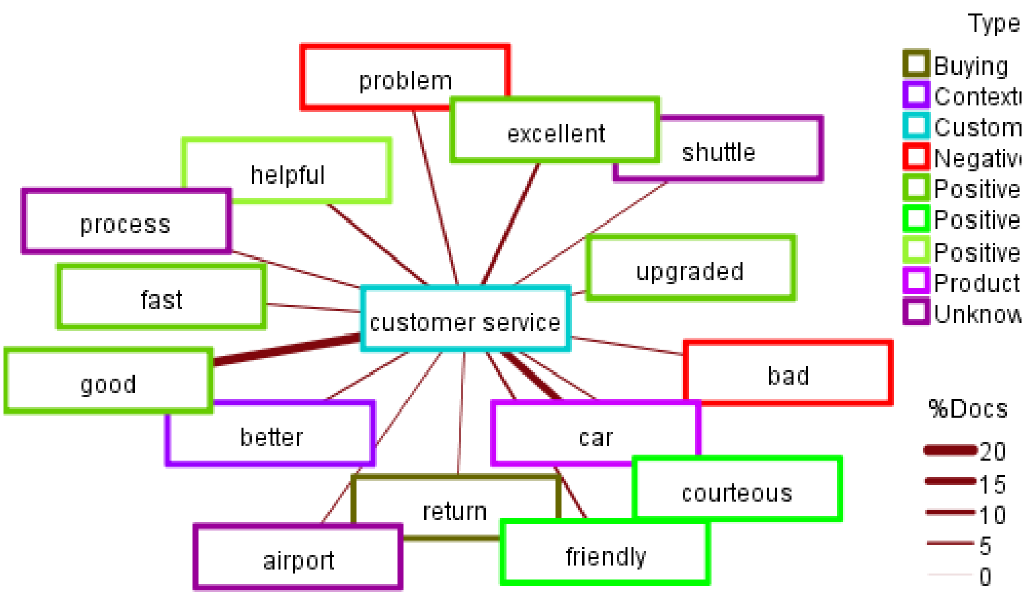
At this point in the analysis, macro level concepts have been analyzed. In addition to this analysis, sentiment analysis that focuses on micro level topics such as customer satisfaction can be helpful for future growth/goal opportunities. While this was briefly explored above from a numerical standpoint, analyzing text in grouped formats is another way to extract sentiment. The text has been color coded to classify the sentiment of specific words. Analyzing how specific text is used in correlation with other terms provides another dimension to text analysis that allows one to analyze the landscape at the ground level. To the left is a chart of all the different types of sentiment found in the documents. Text that is coded in a green hue correlates with positive sentiment, and a red hue is correlated with negative sentiment. The map function in SPSS allows one to see how the different words are associated with one another along with frequency of which the words are used. The color hues allow one to see beyond just the positive or negative, and tap into the context and emotional scale of a word. This analysis was possible for this assignment because the customer satisfaction dictionary was utilized. Customer service is an aspect of the car rental business that is malleable

| Type | In | Global | Docs |
|---|---|---|---|
| <Unknown> | | 225 (23%) | 102 (51%) |
| <Positive> | | 177 (18%) | 128 (64%) |
| <Products> | | 105 (11%) | 80 (40%) |
| <Negative> | | 100 (10%) | 66 (33%) |
| <CustomerSupp | | 98 (10%) | 83 (42%) |
| <PositiveCompe | | 43 (4%) | 39 (20%) |
| <PositiveAttitude | | 38 (4%) | 35 (18%) |
| <Contextual> | | 34 (3%) | 29 (14%) |
| <WaitTime> | | 19 (2%) | 17 (9%) |
| <Budget> | | 19 (2%) | 14 (7%) |
| <Usability> | | 12 (1%) | 12 (6%) |
| <Buying> | | 12 (1%) | 11 (6%) |
| <NegativeAttitud | | 11 (1%) | 11 (6%) |
| <Store> | | 10 (1%) | 8 (4%) |
| <NegativeFuncti | | 10 (1%) | 9 (5%) |
| <Uncertain> | | 10 (1%) | 10 (5%) |
| <Documentation | | 9 (1%) | 9 (5%) |
| <NegativeFeelin | | 9 (1%) | 9 (5%) |
| <PositiveBudget | | 8 (1%) | 8 (4%) |

and directed from a management perspective. Analyzing this text map is a great place to analyze sentiment.

From this map, there are far more positive concepts associated with customer service than negative, in addition the occurrence of the positive concepts used with customer service is greater than the negative sentiment. At first glance, positive customer service is a core competence. Timliness is an areas where customer satisfaction could improve. While terms like excellent, helpful, and good are found linked to the concept ‘too long’, the feedback shows that customers feel like the overall process beginning with the staff and ending with the car simply takes too long. From an operations standpoint, I would impliment a reward system for agents based on timliness. For example, if an agent had a customer through the whole process in less than 15 minutes they would receive a bonus on that specific customers rental.



Through this brief text sentiment analysis, it has been found that there is quite a bit more positive sentiement than negative. Terms like good, friendly, and helpful are most often used to desribe the positive sentiment. The negative sentiment stems from the timliness, and covers the whole rental process. Overall, the car rental agency is doing a lot of things right, and has a couple opportunites to grow. While doing this analysis, I found it particualarly frustrating that SPSS froze and crashed on three separate occasions making the analysis very long. I would like to learn more about the cluster features in the future.