

# ‘Text’-Mining Customer’s View Point and Perceived Value About Brand

Madhumita Ghosh\*

## Abstract

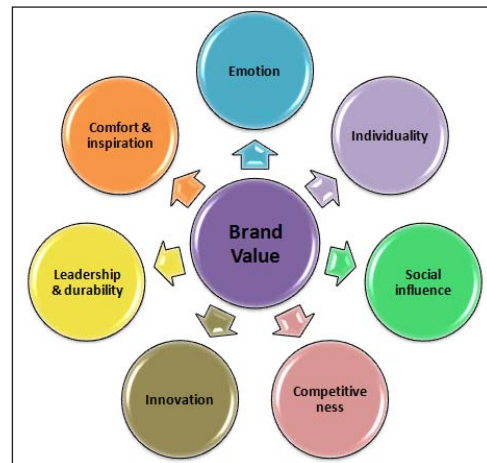
This paper describes how text mining techniques can be applied in the analysis of consumer voice to gain useful and actionable business insights for marketers. The technique is illustrated via its application to understand Brand’s perceived value of certain automobile brands. This case study shows the use of text mining techniques to understand brand’s perception vis-a-vis competition from their opinion, sentiment and reactions. As the amount of online text increases, the demand for text classification to aid the analysis and management of text is increasing. Data acquisition in this case is not costly, information is rich in nature, classification of text can provide this information at low cost, but the classifiers themselves must be built with expensive human effort, or trained from texts which have themselves been manually classified. In this paper, we mention about a procedure of classifying text using the concept of association rule of data mining and correspondence analysis for Brand perception.

Voice of the customer analysis can have significant value for organizations looking to listen to and understand the customer’s “voice” (e.g., from surveys, social media, complaints or web chat) to improve operations and help direct strategy. This approach can, ultimately, help improve customer satisfaction, Net Promoter Score (NPS) and loyalty while reducing churn and dormancy, thus increasing revenues. Consumers’ Experience about a brand depends upon their expectations and engagement across touch-points of the brand. Assess Customer’s Purchase, Usage & Service experience and mindset from social media as emerged touch point helps to understand Brand Imagery

## Introduction

In today’s cut throat competitive marketing, just because a brand wants to stand for greatness doesn’t mean most of the people will convey great things about it. Brands

are not just what they say about them, they are what consumers say they are. A brand’s true identity lies in its ‘perceived value’.



If marketers want to gauge what people think about their brands, there are a variety of (mostly academic) survey methodologies and feedback loops to utilize. Unfortunately, most brand research studies take too long to set-up and administer, and are not timely enough to optimize campaigns. These studies are very expensive and subject to significant survey biases, Social media being a growing touch point between consumer and brands, it is an effective platform to gauge their mindset by mining enormous texts. The honest impression & the most accurate composite of a brand’s true identity seem to come from a consumer’s first gut reaction to it and social sites are the better capture points, where consumers speak their mind.

The qualitative markers go beyond the typical gauges of brand awareness to encompass how consumers feel about a brand, how they think about it, react and talk about it, and interact with it. There are typically Eight trigger points or areas, which is essential to understand in perception studies.

\* Practice Leader - Big Data & Advanced Analysis BA & Strategy - Global Business Services IBM, India.

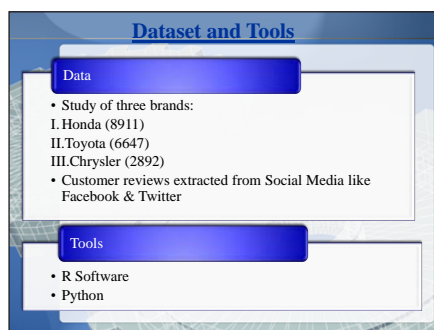
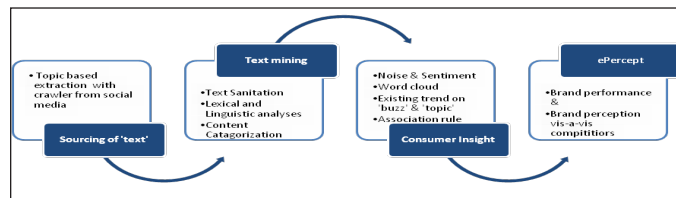
Techniques in text analytics balance existing data mining techniques to help make data richer, and insights more accurate. The technique is built on content management and text search, rather than databases and queries.

### Approach

Keyword searches across various feed often result in the return of too much irrelevant information, which would take a long time to analyze. More sophisticated NLP approach was used, which enables to extract what customers are actually saying about a subject, as opposed to just mentioning a subject, Cluster together the different ways of saying the same thing as well as categorize populations (the text writers) on site according to their behaviours and opinions.

Three Automobile Brands were under study. Viz. Honda, Toyota & Chrysler. Customer reviews were extracted from social media sites with relevant sites and relevant key words.

The basic framework of the complete approach is depicted in this below picture:

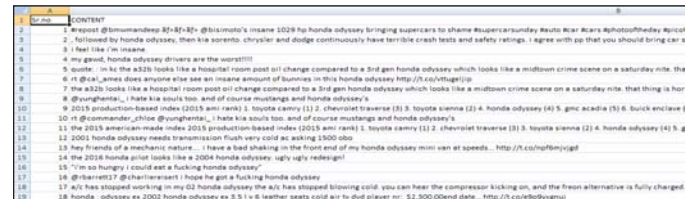


Sourcing of Text: By deploying a propriety tool assess consumer comments from business pages on social media, other industry forums frequented by consumers, and websites pertaining to competitor’s products. The information is collated and checked for any missing or incomplete parts. Connectors were deployed, tested, and fine-tuned to automate the data captured. Various ‘key words’ was tagged to refine the search and to receive more relevant textual information. It helps to minimise the junk at the basic level.

Below picture depicts certain example of texts fetched from social sites.



There were customer reviews for three brands namely Honda, Toyota and Chrysler given by the residents of North America. The corpus of these brands consisted of 8,911, 6,647 and 2,892 related reviews respectively, which was further formatted in database.



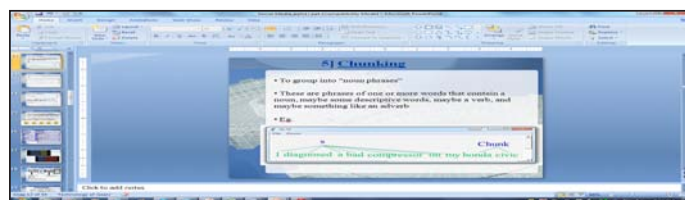
Text Cleansing Process: Majority of available text data is highly unstructured and ‘noisy’ in nature – to achieve better insights and to build better algorithms, it is necessary to clean textual information from various junks e.g. typos, bad grammar, usage of slang, presence of unwanted content like URLs, stopwords, expressions etc. are the usual suspects. The extracted data was cleaned using Natural Language Processing (NLP) technique in Python which involves 7 major steps viz. sentence & word Tokenizing, Removal of Stopwords, Stemming, Parts of Speech Tagging (POS Tagging), Chunking, Chinking and Named Entity Recognition.

Certain Examples:

### Chunking

- To group into “noun phrases”
- These are phrases of one or more words that contain a noun, maybe some descriptive words, maybe a verb, and may be something like an adverb

Example:



### Removing Stopwords

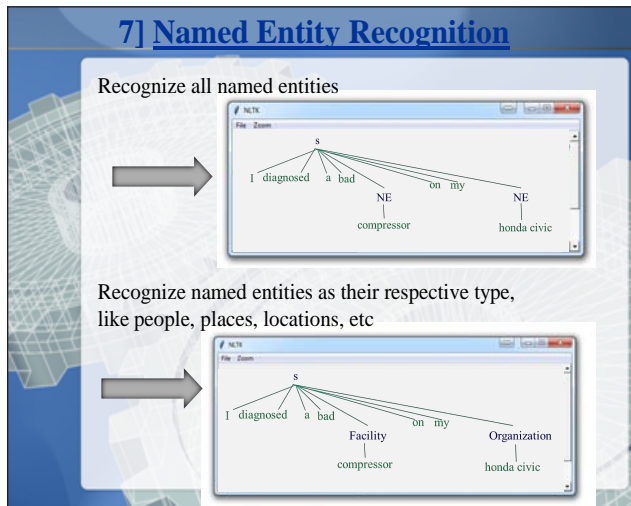
Example:

['T', 'diagnosed', 'a', 'bad', 'compressor', 'on', 'my', 'honda', 'civic', 'yesterday', '.']

Post Stopwords removal:

['diagnosed', 'bad', 'compressor', 'honda', 'civic', 'yesterday', '.']

### Named Entity Recognition



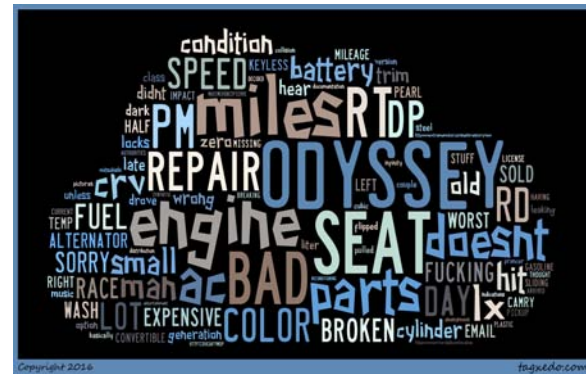
**Text Mining:** The data integration process is concluded with bringing in disparate variables together to create a tabulated data. Text mining, a procedure of synthesizing information, analysing relations, patterns, and rules among textual data. This included word cloud formation, text summarization, text categorization, and text clustering. Text summarization was used to extract partial content reflection. Text categorization on the other hand, assigned a category to the text amongst the predefined categories. A word cloud is a visual representation of text data used to depict keywords.

Larger the word in the visual the more common the word is in the text data. Segregating the comments as per the polarity (negative and positive). As example, two word clouds are shown below for Honda Brand ( for +ve & -ve sentiment) to depict keywords.

Honda Positive Sentiment

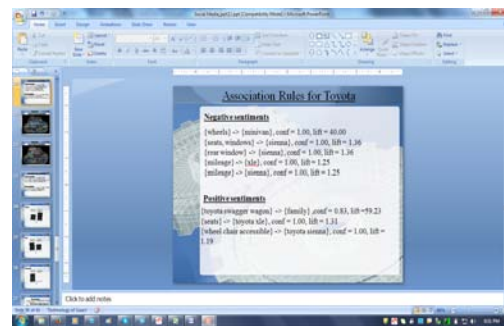


Honda Negative Sentiment



Further with the help of 'Association rule', the content category -sentiment pattern was further strengthened with probability. The rule depict a pattern that states when X (content in text) occurs, Y (sentiment) occurs with certain probability.

Example:



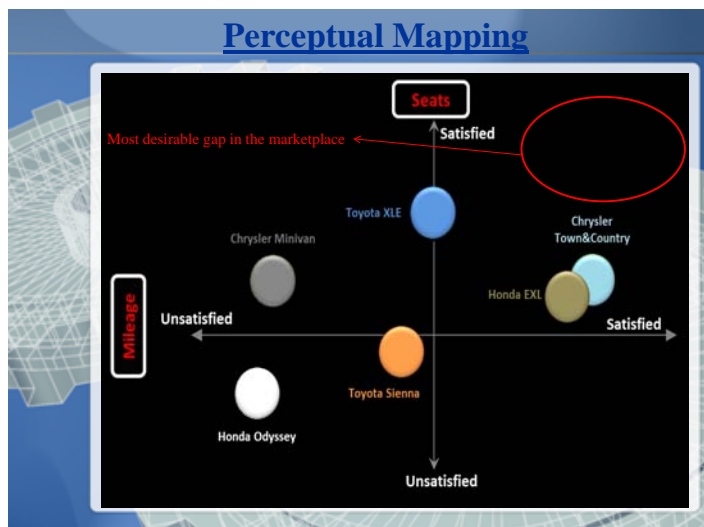
### ePercept

Correspondence analysis, or perceptual mapping, is used

to transform numerical information into a graphical display of a particular market or topic area. Relationships among multiple variables (e.g. brand attributes and brands) are represented in two-dimensional space. Various perceptual distances or proximity between brands and attributes can be compared to gain insight into how various brands are positioned against one another in the minds of the target audience, how each attribute compares to other attributes, and where the brands sit in perceptual space.

We used perceptual mapping to show A particular brand performance vis-à-vis its competitor brands which helped marketer's understand where the brand was lacking as compared to competitor brands.

Output Example with Two Brand attributes 'Seats' & 'Milage'



### Solution Usage & Impact:

Post development of the solution, periodical delivery with 'Brand imagery' (perceptual mapping) of 'The Study Brand' vis a vis competitors which helps gauging customer's expectations and experience.

With appropriate action, The Brand enjoyed an impact of:

- 4% - 8% Improved CSAT score post actions implemented for each Dimensions/Attributes viz. Purchase, Features, Usage & Service.
- Improved Service Management and Brand Advocacy.

### Strategic Action Points:

- Continuous monitoring of brand imagery customer satisfaction, and product feature/service improvement.
- Proactive engagement to enhance customer experience.
- To take focused marketing effort to better position vs. competition and overall market structure.
- To track movement / improvement of themselves and competitive brands.