

Lexicon-based Sentiment Analysis for Product Design and Development

Sutrilastyo^{1*} and Ririn Diar Astanti²

^{1,2} Department of Industrial Engineering, Universitas Atma Jaya Yogyakarta, Yogyakarta, 55281, Indonesia

Email: 195603118@students.uajy.ac.id, ririn.astanti@uajy.ac.id

*Corresponding author

ABSTRACT

This paper discusses the role of text mining and sentiment analysis in collecting and analyzing customers' verbatim/voice of the customer for product design and development. In the illustration case of designing car underbody, the data were collected from a car online forum discussion website and processed using text mining techniques with "underbody" as a keyword. The result of the analysis finds there are worries regarding underbody durability against rust and damages. This finding is used as a reference point for the car underbody design process.

DOI: <https://doi.org/10.24002/ijieem.v3i1.4351>

Keywords: Sentiment analysis, product design, product development, the voice of the customer.

Research Type: Research Paper

Reference to this paper should be made as follows: Sutrilastyo & Astanti, R.D. (2021). *Lexicon-based Sentiment Analysis for Product Design and Development*. *International Journal of Industrial Engineering and Engineering Management*, 3(1), 29–33.

1. INTRODUCTION

The first thing to be addressed in product development is identifying customers' expectations on the would-be product. Customers' expectations and the product's derived value are important aspects of quality (Ikiz & Özdağoğlu, 2015; Mulay & Khanna, 2017). An appealing product has an impact on image formation and positive brand and product evaluations for customers (Kreuzbauer & Malter, 2005).

Customer expectations, often called Voice of Customer (VoC), are collected through surveys, interviews, focus group discussion, dedicated feedback form, and other tools. However, these methods face several setbacks in most situations, such as cost efficiency, time, and place limitations. To overcome these limitations, current technology allows many platforms for customers to share their views, thoughts, and experience through social media, product reviews, blogs, and forum discussions. These large data sets are valuable to enable researchers to find hidden patterns in customers' wants and needs.

Text mining is a method to extract knowledge from texts (Khedr et al., 2017). Text mining enables the information of what customers expect of a product to be extracted without going through lengthy and complicated

methods when facing large sets of data (Ikiz & Özdağoğlu, 2015). Sentiment analysis is one of the text mining methods to determine the contextual polarity of writing, whether it expresses negative, neutral, or positive opinions (Shukri et al., 2015). The method explores customers' views, thoughts, and experiences on one or more keywords related to the product and helps frame customers' expectations.

This paper discusses the role of text mining and sentiment analysis in product design and development, especially its role in the idea screening stage. This concept is illustrated in the application of car underbody design. Previous research on underbody design mainly focuses on its technical performance, such as improving its aerodynamics (Buljac et al., 2016; Desai et al., 2008; Heidemann et al., 2018; Yuan & Wang, 2017). Other researchers also discuss tackling issues in the underbody such as squeak or using better materials to improve performance and tailor better to customer's specifications (Dittmar & Plaggenborg, 2019; Moos, 2014; H. Park, 2019). To the best of the Author's knowledge, there is no research on underbody design that specifically considers customers' wants in an underbody. This study tries to fill that gap by using text mining to collect VoC in underbody design to create an appealing product for customers.

2. LITERATURE REVIEW

2.1. Product development

In developing a new product, Kotler & Keller (2006) proposed eight steps of new product development, namely idea generation, where the idea is formulated through personnel brainstorming. The generated ideas are selected and narrowed down in the idea screening step. The concept development and testing step define a product's parameters and measures customers' purchase intention based on the selected idea. Detailed strategies are developed and analyzed through marketing strategy and business analysis steps. Product development, test marketing, and commercialization steps focus on the production of new products, testing new products in a trial market, and launching the new product. Finding customers' expectations of a product through VoC gives the idea generation process input.

Customers' verbatim is an essential resource for companies as they include customers' expectations and requirements and help companies improve their products or services (Aguwa et al., 2017). Text mining helps to remove the labor-intensive and time-consuming process of surveys, interviews, and other commonly-used data collecting for VoC. Their study on text mining as a supporting process for VoC clarification found that text mining techniques help to hasten VoC processes (Ikiz & Özdağoğlu, 2015).

2.2. Sentiment analysis

There are three levels where sentiment classification can be performed: document-level, sentence-level, and aspect and feature level (Singh & Sharma, 2013). As the name implies, document-level sentiment classification determines the polarity of an entire document and sentence-level for each sentence. Aspect and feature level sentiment classification classifies documents or sentences according to certain aspects of sentences.

In general, sentiment classifications are done by doing the following actions in order: document pre-processing, feature selection, feature extraction, and text classification (Shah & Patel, 2016). Feature extraction is a process of extracting new features based on the features generated in the feature selection phase. It can be done with unigram-based, bigram-based, unigrams with bigrams, and unigrams with POS (Abdulla et al., 2013; Alsaeedi & Khan, 2019; Shah & Patel, 2016). In unigram-based sentiment classification, the basic approach is to collect sentiment scores or labels associated with the unigram from a designated lexicon with an assigned polarity score (Dey et al., 2018).

The use of sentiment analysis in the idea generation step of product design and development has been discussed in previous studies. Jeong & Yoon (2016) used sentiment analysis to identify product development opportunities. Product features with high opportunity were able to be identified and used to formulate product development. Using sentiment analysis on feature-based review data, J. Park et al. (2018) identified features that should be given more importance in the improvement of tires. These

studies show sentiment analysis's role in the idea generation step of new product development, either developing new products or improving existing products according to customers' expectations and requirements.

2.3. Underbody design

Underbody design is of great interest in the automotive industry (Desai et al., 2008). One of the reasons for this interest is related to the decrease in fuel consumption. Improved fuel economy could be achieved with an underbody design that aerodynamically reduces total drag force (Heidemann et al., 2018; Muyl et al., 2004; Yuan & Wang, 2017). Humenic & Humenic (2020) found that the use of curved diffusers in the underbody leads to small values of drag, and from a design point of view, it has little interference with other components of cars.

In their research of drag reduction device of a trailer underbody and base, Ortega & Salari (2004) found that curved base flaps were the most effective drag-reduction add-on device for a trailer. Meanwhile, Cho et al. (2017) did similar research on the sedan and found a combination of an undercover and side air dam had the best result in reducing aerodynamic drag. Synergy effect was obtained when combining the two devices resulting in the maximization of undercover's diffusing effect, thus increasing rear surface pressure and reduction in the wake region.

3. PROPOSED FRAMEWORK

This study is trying to explore the use of sentiment analysis to discover customers' expectations and requirements of the product. The requirements are used as guidelines to determine the direction of the idea in the first step of the product development phase, hence fulfilling customer expectations. The proposed method of this research is shown in Figure 1. The data are collected using the web scraping method on social media, online discussion forums, blogs, or product reviews to collect the customers' thoughts verbatim.

After the data are collected, common text pre-processing techniques, i.e., data cleaning, tokenization, stemming, and removing stop words, are used to clean the data. Data cleaning is a process to remove unwanted characters or gibberish from the text. Tokenization decomposes sentences into words or phrases called tokens (Ikiz & Özdağoğlu, 2015). Each token is returned to its root words by removing suffixes in the stemming process. Commonly used words (stop words) are removed to give more emphasis on words that define the meaning of the text. This process leaves only nouns and adjectives as tokens.

Once the data were pre-processed, advanced analysis techniques can be deployed for further inferences, i.e., assigning sentiments to each token using the sentiment lexicon package from Hu & Liu (2004). The labeled unigrams/tokens are ranked based on the frequency of the word, and they are used as the basis of idea generation.

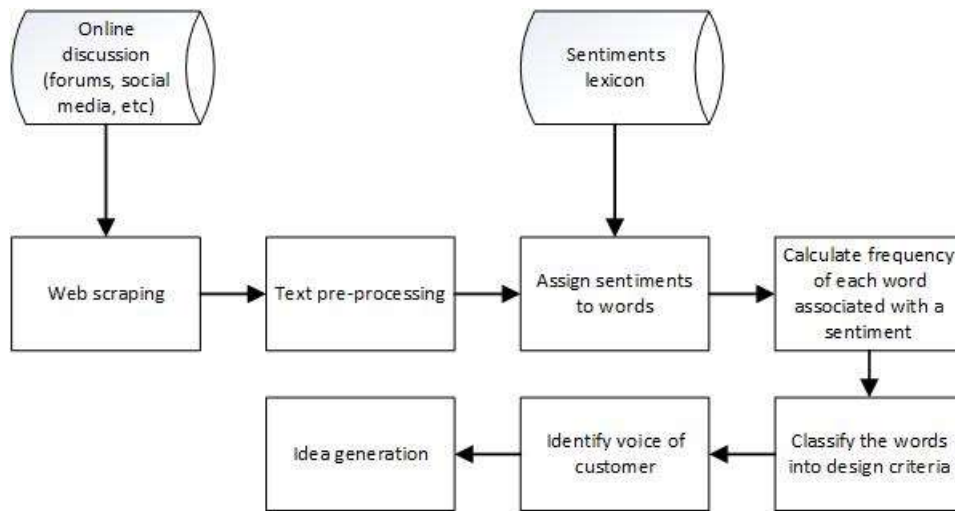


Figure 1. Proposed framework

Table 1. Summary of website selection

Website	Users	Car-specific	Active Discussion	Region-specific
Facebook	X			
Twitter	X			
Kaskus (Otomotif subforum)		X		X
My Car Forum		X	X	X
Reddit (r/Cars subforum)	X	X	X	

Table 2. Example of an extracted data frame

...	Author	User	Comment score	Comment	...
...	Anonymous	Awanama	3	I love that car	...
...	Anonymous	Nirmala	1	The suspension was kinda bad	...
...

4. ILLUSTRATED CASE STUD

The application of the proposed framework was illustrated in the design process of the car underbody. Several social media and forums were considered as data sources. The initial selection of social media and forums was based on Author’s on-hand knowledge about the most popular social media and forum: Facebook, Twitter, and Reddit. The other forums were found using Google search with “car forum” or “forum mobil” as keywords.

The next step was to determine the criteria for website selection. The Author determined four criteria for website selection: the platform has to have more than one million users, specifically discusses cars, has an active discussion about car-related topics, and is not region-specific. The result of the selection is summarized in Table 1.

From the criteria selection table above, Reddit, specifically the r/Cars subforum, was chosen as the data source. Cen (2020) stated that Reddit offers better external validity compared to other niche web forms. One of the criteria used in scrapping Reddit data from Reddit was that the threads should at least have fifty comments to ensure the scraped data were from active threads.

The data were extracted with Rstudio using “underbody” as a keyword to search all threads containing the word. The extracted data were in a data frame illustrated in Table 2.

5. RESULT AND DISCUSSION

The use of sentiment analysis in the idea generation step of product development is illustrated with the process to design a car underbody. Gurusami (1999) stated there are fifteen criteria that need to be considered in designing underbody: investment, variable cost, timing, weight, customer satisfaction, NVH (noise, vibration, harshness), vehicle dynamics, rear crash performance, craftsmanship, durability, fuel tank, seating location, rear overhang, types of rear suspension, types of exhaust.

The result of sentiment analysis of keyword “underbody” in Figure 2 shows that “rust” is the most mentioned negative sentiment word. This result is in line with Rana et al.’s (2018) finding that underbody coating is prone to corrosion. Moreover, other words such as “damage” and “issues” are words related to the aforementioned durability criteria. We can see that although customers verbatim did not mention durability explicitly, we can infer these words fit into durability

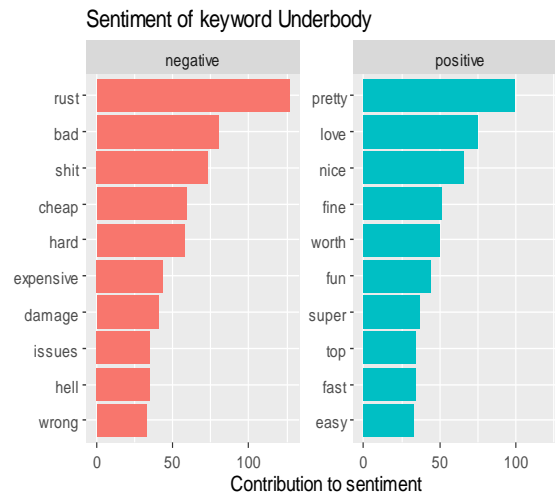


Figure 2. The sentiment of the “underbody” keyword

criteria. On the other hand, “pretty” is the most mentioned positive sentiment word that we can infer will fit into aesthetics criteria that are not included in the aforementioned fifteen criteria. This finding shows that by scraping an online discussion forum and processing them using text mining, we got a voice of the customer that focuses more on underbody aesthetics aspects. Previous research mainly focuses on manufacturing and technicalities of underbody design. The proposed framework complements the existing design process by bringing the technical aspects and what the customers want together. By capturing customers’ perceptions through specific words, companies are able to crosscheck whether their definition of criteria is aligned with customers’ definitions of criteria.

This finding is valuable in product design and development. It can be used as a guideline in narrowing down the options and directing the direction the product should take to meet customers’ expectations. Although customer expectations can be fulfilled in a later phase by tailoring the specifications according to customers’ wants (Moos, 2014), integrating the expectations in the idea generation phase will incur less cost of change (Folkestad & Johnson, 2001). In the idea generation phase of the illustrated case, the ideas focusing on underbody durability and aesthetics should be prioritized.

6. CONCLUSION

This research explores the role of sentiment analysis in the idea screening stage of product development. Text mining demonstrates an advantage time and cost-wise in VoC data collection and analysis. In the illustrated case discussed above, sentiment analysis contributes to the idea-generating stage of product development, specifically in narrowing down the options and determining the product’s direction. This research also finds that sentiment analysis is able to verify companies’ perspectives of criteria with customers’ perspectives.

From Author’s findings, there is a gap in the Author’s knowledge around sentiments analysis in a car’s underbody design process, namely the ambiguity of one-word sentiments. Further research

on the use of n-grams in doing the sentiment analysis is hypothesized to be able to reduce this ambiguity.

REFERENCES

- Abdulla, N. A., Ahmed, N. A., Shehab, M. A., & Al-Ayyoub, M. (2013). Arabic sentiment analysis: Lexicon-based and corpus-based. *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2013*.
- Aguwa, C., Olya, M. H., & Monplaisir, L. (2017). Modeling of fuzzy-based voice of customer for business decision analytics. *Knowledge-Based Systems, 125*(November), 136–145.
- Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications, 10*(2), 361–374.
- Buljac, A., Kozmar, H., & Džijan, I. (2016). Aerodynamic performance of the underbody and wings of an open-wheel race car. *Transactions of Famena, 40*(2), 19–34.
- Cen, P. (2020). *Predicting Consumers’ Brand Sentiment Using Text Analysis on Reddit* [Joseph Wharton Scholars]. https://repository.upenn.edu/joseph_wharton_scholars/95.
- Cho, J., Kim, T.-K., Kim, K.-H., & Yee, K. (2017). Comparative Investigation On The Aerodynamic Effects Of Combined Use of Underbody Drag Reduction Devices Applied to Real Sedan. *International Journal of Automotive Technology, 18*, 293–300.
- Desai, S., Leylek, E., Lo, C. M. B., Doddegowda, P., Bychkovsky, A., & George, A. R. (2008). Experimental and CFD comparative case studies of aerodynamics of race car wings, underbodies with wheels, and motorcycle flows. *SAE Technical Papers, February 2018*.

- Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, 103, 92–105.
- Dittmar, H., & Plaggenborg, H. (2019). Lightweight vehicle underbody design. *Reinforced Plastics*, 63(1), 29–32.
- Folkestad, J. E., & Johnson, R. L. (2001). Resolving the conflict between design and manufacturing: Integrated Rapid Prototyping and Rapid Tooling (IRPRT). *Journal of Industrial Technology*, 17(4), 1–7.
- Gurusami, S. A. (1999). Decision analysis in product development - A case study. *SAE Technical Papers*, 724.
- Heidemann, R., Rodrigues, A. F. A., Bohrer, A., Gertz, C. L., & Cervieri, A. (2018). Underbody aerodynamics: Drag coefficient reduction in road vehicles. *SAE Technical Papers*, September.
- Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. *Proceeding of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>.
- Huminic, A., & Huminic, G. (2020). Aerodynamics of curved underbody diffusers using CFD. *Journal of Wind Engineering and Industrial Aerodynamics*, 205(July), 104300.
- Jeong, B., & Yoon, J. (2016). *Identifying product opportunities using topic modeling and sentiment analysis of social media data*. <http://apiems2016.conf.tw/site/userdata/1087/papers/0208.pdf>.
- Ikiz, A. K., & Özdağoğlu, G. (2015). Text Mining as a Supporting Process for VoC Clarification. *Alphanumeric Journal*, 3(1), 25–40.
- Khedr, A. E., Salama, S. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, 9(7), 22–30.
- Kotler, P., & Keller, K. L. (2006). *Marketing Management* (12 Edition). Prentice-Hall.
- Kreuzbauer, R., & Malter, A. J. (2005). Embodied cognition and new product design: Changing product form to influence brand categorization. *Journal of Product Innovation Management*, 22(2), 165–176.
- Moos, E. (2014, April 1). *Light Weight Materials for Aerodynamic, Acoustic and Thermal Management Improvements in Underbody Applications*.
- Mulay, R., & Khanna, V. T. (2017). A Study on the Relationship between the Voice of Customer with the Cost of Quality in Processes of Professional Higher Education Institutions. *South Asian Journal of Management*, 24(4), 55–72.
- Muyl, F., Dumas, L., & Herbert, V. (2004). Hybrid method for aerodynamic shape optimization in automotive industry. *Computers and Fluids*, 33(5–6), 849–858.
- Ortega, J. M., & Salari, K. (2004). An experimental study of drag reduction devices for a trailer underbody and base. *34th AIAA Fluid Dynamics Conference and Exhibit*.
- Park, H. (2019). *Development of a Test System to Measure Squeak Propensity of Vehicle Underbody Components*. University of Cincinnati.
- Park, J., Lee, H., Lee, J. H., & Suh, H. (2018). Feature-based sentiment word selection and rating for system design. *Journal of Industrial Electronics Technology and Application*, 1(4), 54–57.
- Rana, V. S., Gosai, D. C., & Joshi, D. S. (2018). Enhancement of Car Life, Engine Performance, and Emission Test by Using Coatings. *International Journal of Modern Trends in Engineering and Science*, 05(2017), 44–49.
- Shah, F. P., & Patel, V. (2016). A review on feature selection and feature extraction for text classification. *Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2016*, 2264–2268.
- Shukri, S. E., Yaghi, R. I., Aljarah, I., & Alsawalqah, H. (2015). Twitter sentiment analysis: A case study in the automotive industry. *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2015*.
- Singh, S. R., & Sharma, S. (2013). A global optimizing policy for decaying items with ramp-type demand rate under two-level trade credit financing taking account of preservation technology. *Advances in Decision Sciences*, 2013, 1–16.
- Yuan, Z., & Wang, Y. (2017). Effect of underbody structure on aerodynamic drag and optimization. *Journal of Measurements in Engineering*, 5(3), 194–204.

This page is intentionally left blank
