Integrating Sentiment Analysis and Quality Function Deployment for Product Development

Abdullah 'Azzam [a,1], Syafira Mahardiningtyas [a,2], Qurtubi [a,3]
[a] Departement of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia, Kaliurang St Km.14 Sleman, Yogyakarta Indonesia
[1] abdullah.azzam@uii.ac.id; [2] 16522242@alumni.uii.ac.id; [3] qurtubi@uii.ac.id[*]
[*] corresponding author

ARTICLE INFO

Article history
Received 27 Sept 2023
Revised 18 Oct 2023
Accepted 29 Nov 2023

Keywords
QFD
Machine learning
Naïve Bayes
SVM

ABSTRACT

The development of technology and media has made online data reviews a promising data source. Through machine learning utilizing text processing, data analysis of Ventela Public Low product reviews can be carried out—sentiment analysis is used to find class groups from each data. The classification algorithm is Naïve Bayes and Support Vector Machine (SVM). A classification model with the best performance and accuracy values will be selected. Word association is then applied to obtain information from the required class. Quality Function Deployment (QFD) is a tool used to assist designers in developing products. The results of the integration of sentiment analysis into QFD show that sentiment analysis produces information by the provisions of the QFD method and can support the product development process in terms of the amount of data various data topics and reduces the subjectivity of designers at the stage of determining voice of customer (VOC) and performance values of products and competitors.

1. Introduction

The use of the internet has become a part of everyday life for the majority of people in the world. Internet World Stats in 2017 revealed more than 4 billion internet users worldwide. Of the 4 billion internet users, 3 billion are active social network users. From the internet, Google's site became the most visited site, amounting to 16.38%, then Facebook at 5.89%, YouTube at 2.94%, and the accumulative of other sites by 74.79%. From the accumulative sites that exist, Amazon is the site that reaches the fifth position most visited by the public. Many internet users consider digital media as a means to shop for the desired and needed products or services [1].

The ease of sharing information is a great opportunity and a challenge for those who can use it. The opinions or opinions of others can influence the decision-making process. At this time, a person's opinion can be accessed from several sources, including through online review/review websites, personal blogs, and social media. Reviews are opinions of someone, not an advertisement, and are one of the factors that determine decisions regarding a person's purchasing power [2]. Reviews are included in big data, which is a developing term that describes volume in a structured, semi-structured, and unstructured manner in a data set that has the potential to be processed into information [3].

In this study, a data analysis model with a machine learning approach and QFD will be developed using online review data for products made in Indonesia, namely Ventela Public Low shoes. Natural language processing (NLP) is one of the fields of Artificial Intelligence that studies communication between humans and computers through natural language. Sentiment analysis will be carried out to map data with algorithms popular in Indonesian language data processing and have good performance, namely Naïve Bayes and SVM. The selection of Ventela products is due to the high interest of Indonesian consumers in these products, as evidenced by the large number of reviews available on e-commerce and social media. Over the last four years, the domestic shoe industry has increased, with the emergence of various new brands [4], and Ventela is one of the most popular brands. The number of shoes sold is estimated at 700 to 1000 pairs per month [5]. Many products fall into the category of failure, one of which is because of using assumptions. An example case is the Windows Phone, which cannot beat the domination of markets [6]. The results obtained from this study are expected to help designers make a product and meet consumer needs with more real-time data and, in the end, can increase the chances of the product's success in the market.

Data analysis through sentiment is a growing area of research. The purpose of sentiment analysis is to find opinions/opinions in the writing of the text, get the sentiments that exist in these opinions, and finally get the polarity/classification of these sentiments, whether positive/supportive or negative/unsupportive. Often used algorithms are divided into two, namely dictionary-based and machine learning-based algorithms. SVM and Naïve Bayes are algorithms commonly used in sentiment analysis with data used in Indonesian. Based on several studies that have been carried out, the classification algorithm can improve its performance, one of which is mutual information, which acts as a feature selection.

This study will develop an online review data analysis model with a mutual information approach as feature selection, TFIDF, and Bags of Words (BoW) as word weighting, then SVM and Naïve as classifiers. With these approaches, experiments will be conducted to find the best-performing model and hopefully improve the sentiment analysis process on online review data. Furthermore, to complete the product development process about product design, the results of the sentiment analysis will serve as the VOC, which is used as input in the Quality Function Deployment (QFD) method. The integration of data analysis using machine learning with the QFD method is proposed by researchers to help business owners in the product design process by optimizing existing data and with better processes.

2. Method

2.1. Sentiment analysis

First, The social media used in data collection are Twitter and YouTube, and the marketplace platform used is Shopee. The selection of social media and marketplace platforms is based on a questionnaire conducted with Ventela user respondents and the ease of collecting data on these social media and platforms. The total data obtained by Ventela Public Low products is 10412 data with 1673 usable data and Converse products, from the total number of reviews, as many as 14346, the number of appropriate data to be used as a comparison is 469 data reviews. The data has been added gradually, document A (1062 data), document B (1176 data), and document C (1326 data). Each document has the same positive, negative, and neutral class proportions. The algorithms used for sentiment classification are Naïve Bayes (Multinomialnb) and SVM (SVC parameter C = 1.0, kernel = 'linear', gamma = 'scale'), as algorithms that are widely used in Indonesian language data and have good performance. The word weighting algorithm used is Term Frequency Inverse Document Frequency (TFIDF) and Bag of Word (BoW); the random number used is 77. Mutual Information (MI) and TFIDF L1 are the feature selection algorithms used.

The first process is manual labeling, which provides an initial identity to the supervised learning algorithm. Creating automated text categorization constructs requires knowledge from a domain specialist [7]. This is useful for comparing system analysis with linguist analysis. However, in this study, due to time and cost limitations, the initial labeling process was not carried out by experts but more than once, namely by five people with different educational backgrounds.

In optimizing the model, a series of experiments was carried out. The confusion matrix will describe the results of the correct positive prediction accuracy, false positive predictions, correct negative predictions, wrong negative predictions, correct neutral predictions, and wrong neutral predictions. Accuracy calculations based on all correct predictions are compared with all test data. The higher the accuracy value, the better the resulting model. Performance assessment is measured based on the precision, recall, and f-measure performance values. A model is said to have a good value if both precision and recall have high values; this shows that the model is free from bias [8].

2.2. QFD

QFD is a method of planning and developing a product or service in a structured manner [9]. The QFD method provides several benefits to organizations trying to increase their competitiveness by improving their quality and productivity. Some benefits include customer focus, time efficiency, teamwork orientation, and documentation orientation [10]. QFD is a structured process or mechanism to determine customer needs and translate these needs into relevant technical requirements, where each functional area and organizational level can understand and act [11].

3. Results and Discussion

3.1. Classification Model Validation

Evaluation serves to determine the accuracy of the proposed algorithm model. Validation is used to compare the results of the accuracy of the model used with the results that have been there before. The validation technique used is 10-fold Cross-Validation. During validation, the order of the existing documents will be randomized. This aims to avoid grouping documents that come from specific categories. The results of each test can be seen in Table 1 below.

Table 1. The results of the 10-cross validation of the selected model

K		Precision	Recall	F1-score	Support	K		Precision	Recall	F1-score	Support
1	negative	0,56	0,53	0,55	34	6	negative	0,92	0,83	0,88	42
	neutral	0,51	0,68	0,58	31		neutral	0,87	0,87	0,87	47
	positive	0,93	0,79	0,86	53		positive	0,85	0,97	0,9	29
	accuracy			0,69	118		*accuracy*			0,88	118
2	negative	0,71	0,63	0,67	43	7	negative	0,82	0,89	0,86	37
	neutral	0,62	0,69	0,65	35		neutral	0,93	0,83	0,88	47
	positive	0,88	0,9	0,89	40		positive	0,94	1	0,97	33
	accuracy			0,74	118		*accuracy*			0,9	117
3	negative	0,82	0,89	0,85	46	8	negative	0,79	0,79	0,79	39
	neutral	0,83	0,69	0,76	36		neutral	0,81	0,83	0,82	36
	positive	0,89	0,94	0,92	36		positive	0,95	0,93	0,94	42
	accuracy			0,85	118		*accuracy*			0,85	117
4	negative	0,81	0,91	0,86	33	9	negative	0,86	0,97	0,91	33
	neutral	0,94	0,81	0,87	37		neutral	0,98	0,86	0,92	51
	positive	0,98	1	0,99	48		positive	0,91	0,97	0,94	33
	accuracy			0,92	118		*accuracy*			0,92	117
5	negative	0,84	0,86	0,85	43	10	negative	0,84	0,88	0,86	42
	neutral	0,87	0,85	0,86	40		neutral	0,84	0,84	0,84	32
	positive	0,97	0,97	0,97	35		positive	0,93	0,88	0,9	43
	accuracy			0,89	118		*accuracy*			0,87	117
				average accuracy			*0,850463*				

The resulting accuracy value for each fold has varied results. The difference in accuracy for each fold is due to the different characteristics of the test data and training data on each fold. Folds 4 and 9 produce the highest accuracy (0.92) because the test data and training data have similar characteristics, while in fold-1, which has the lowest accuracy (0.69), the training data and test data have quite a few characteristics. The average accuracy obtained is 0.850463.

In addition, algorithm evaluation is also carried out with the ROC (Receiver Operating Characteristic) curve. The ROC curve can be used to get the AUC value. The AUC value is used to determine the accuracy of the diagnostic test classification. There are several levels of diagnostic value as follows [12]:

a. Accuracy 0,90-1,00 = excellent classification

b. Accuracy 0,80-0,90 = good classification

c. Accuracy 0,70-0,80 = fair classification

d. Accuracy 0,60-0,70 = poor classification

e. Accuracy 0,50-0,60 = failure classification

Based on the results obtained, the AUC value model accuracy is at the level of good classification. The results of the ROC-AUC value are as shown in figure 1.

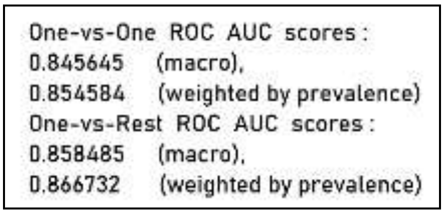
One-vs-One ROC AUC scores:
0.845645 (macro),
0.854584 (weighted by prevalence)
One-vs-Rest ROC AUC scores:
0.858485 (macro),
0.866732 (weighted by prevalence)

Fig. 1.ROC-AUC results for selected models

3.2. Prediction and word association

The review data, amounting to 1673 with a manual labeling process, was divided into positive classes with 746 data, negative as many as 476, and neutral as many as 451 data. Meanwhile, through the prediction process with the selected model, the results obtained were 602 positive class data, 547 negative class data, and 524 neutral data. Through the evaluation process, 475 predictive data do not match the results of manual labeling, which means the accuracy of the prediction process with 1673 review data is 71.60789.

The standard procedure for evaluating association measures uses a manual assessment of the n-best candidates identified [13]. The proposed manual assessment is to rate true positive (tp); in this case, it relates to the reasonable terms in the 50 or 100 terms that have the highest ranking. Perform comparisons and evaluations using log-likelihood, t-score, chi-squared, MI, and frequency methods manually at different n-best, and it is found that n = 100 has the best precision values results among the other best n-best [14]. So, in this study, the selection of the association method will be based on true positives at a rating of 1-100 for each method.

The results of the Ventela product association using the selected method, namely the t-test, have the highest value of 2.22860361 with terms (make, design, yourself) and the lowest value of 0.906567235 with terms (make, brand, not). The difference between the top and lowest terms is 1.322036. In Converse products, the best association method in rank 1-100 is the frequency method, which has the highest value of 3 with terms (more, buy, vans) and (more, more, more). Meanwhile, the lowest value is one and is owned by many terms, indicating that the data is very varied.

3.3. VoC Identification

Identification of features of consumer needs is carried out using data from Ventela product reviews, ranking 1-100 of the prediction results associated with the previously chosen method. The Converse product review data will be used as a benchmark product, ranking 1-100 of the results with the same process stages as those applied to the Ventela product review data.

The following process is to match the features of Ventela's consumer needs with competing products—selection of competitor product data on each data based on the similarity of categories with the selected data. The data used is the highest value in that category because competitor products are used as a comparison. Table 2 contains the results of the previous process stages used in the QFD stage. The results of data that have gone through the matching process can be seen in Table 3.

Table 2. Example of word association results as QFD input

Ventela		Converse	
Trigram	t	Trigram	Freq
(make, design, own)	2,22860361	(over, buy, vans)	3
(no, as, compass)	1,99212736	(more, mold, more)	3
(characteristic, typical, own)	1,732023663	(price, sale, crazy)	2
(vans, you, too)	1,731879571	(ventela, brand, local)	2
(love,'product, indonesia)	1,731700185	(measuring, no, want)	2

Table 3. The process of matching each data

Ventela	Converse

Trigram	t	Trigram	Freq
(make, design, own)	2,22860361	(more, mold, more)	3
(no, as, compass)	1,99212736	(more, mold, more)	3
(characteristic, typical, own)	1,732023663	(more, mold, more)	3
(vans, you, too)	1,731879571	(over, buy, vans)	3
(love, product, indonesia)	1,731700185	(ventela, brand, local)	2

Data that have gone through the matching process are then grouped. The classification applies to both Ventela products and competitors' products. In Ventela products, the grouping is based on the existence of the same term, while for competitor products, the grouping is done by following the results of the Ventela grouping. The grouping process can be seen in Table 4.

Table 4. Example of data grouping process

	Ventela		Converse	
	Trigram	t	Trigram	Freq
Group1	(make, design, own)	2,22860361	(more, mold, more)	3
	(have, design, own)	1,730218524	(more, mold, more)	3
	(need, make, design)	1,413071414	(more, mold, more)	3
	(product, indonesia, design)	1,411637026	(more, mold, more)	3
	(search, make, design)	1,406218523	(more, mold, more)	3
	('make, design, new)	1,404314943	(more, mold, more)	3

3.4. Analysis of product and service performance levels

The The result of the association process of each data has a value that represents the closeness between words and how often these terms co-occur. This value will be used to calculate the level of performance of products and competitors, given the similarity of the concept; namely, the negative data class that appears most often or is most often discussed can have poor performance. Calculation of the level of product performance begins with the normalization of each value. From the negative class, it has a minimum value of 2,228604 and a maximum value of 1,400317. Each grouping result will then calculate the normalized average of the forming data values. The final product performance results for each feature are obtained from the average product performance multiplied by the weight of each feature.

The same calculation steps are carried out in competitor products, but there are differences because the competitor products used are all classes. So, for the normalization process, the minimum value is 1, and the maximum is 3. The results of product performance levels and competitors are shown in Table 5 below.

Table 5. Product value and competitor performance

Group	Ventela	Converse	*Result*	Group	Ventela	Converse	*Result*
Group1	0,0457	0,06	*bad*	Group13	0,0394	0,02	*bad*
Group2	0,0319	0,04	*bad*	Group14	0,0493	0,025	*bad*
Group3	0,0356	0,03	*bad*	Group15	0,0197	0,01	*bad*
Group4	0,0948	0,05	*bad*	Group16	0,0915	0,075	*bad*
Group5	0,0159	0,01	*bad*	Group17	0,0197	0,01	*good*
Group6	0,0792	0,04	*bad*	Group18	0,0197	0,02	*bad*
Group7	0,095	0,075	*bad*	Group19	0,0099	0,01	*bad*
Group8	0,0198	0,01	*bad*	Group20	0,0099	0,01	*bad*
Group9	0,1446	0,075	*good*	Group21	0,0099	0,005	*bad*
Group10	0,0258	0,0201	*good*	Group22	0,0099	0,005	*bad*
Group11	0,0357	0,04	*bad*	Group23	0,0099	0,005	*bad*
Group12	0,0029	0,01	*bad*	Group24	0,0099	0,01	*bad*

3.5. Determination of Important Rating

The feature formation process will carry out the results of the grouping. Features are tailored to the characteristics of Ventela's products and consider the associated value of competitors' products. If the product performance value is lower than the competitor's product performance value, the features are negative, and vice versa. If the product performance value is better than the competitor's performance, the features are positive, as can be seen in Table 6.

The next stage measures the level of important rating (IR) on the features obtained. In contrast to the level of product performance that can use other available data, at this stage, measuring the importance of features cannot be done by inputting the associated value because the level of importance of a feature is better than doing a direct assessment of the feature [15]. So, in this study, the IR value was obtained from the results of a questionnaire that was distributed to 55 respondents who used Ventela products. To equalize the size, the normalization process will be carried out on the IR results, with a maximum value of 5 and a minimum of 1. Table 7. contains the IR value of each feature.

Table 6. Example of feature formation

Initial	Feature	Ventela	Converse	Result
F1	Product design	0,0457	0,06	bad
F2	The level of originality of the product	0,0319	0,04	bad
F3	Similarities to products on the market	0,0356	0,03	bad
F4	Size availability needs to be increased	0,0948	0,05	bad
F5	Change the sizing	0,0159	0,01	bad
F6	Make large sizes	0,0792	0,04	bad
F7	Imitating Vans	0,095	0,075	bad
F8	Imitating *Converse*	0,0198	0,01	bad
F9	Proud of local product	0,1446	0,075	good
F10	Indonesian Products	0,0258	0,0201	good
F11	Product Model	0,0357	0,04	bad
F12	Better than Compass	0,0029	0,01	bad

Table 7. Example of IR value of each feature

Feature	Average of IR	Normalization	IR(roundup)	Ranking
F1	4,709091	0,927273	0,9273	2
F2	4,563636	0,890909	0,891	4
F3	3,654545	0,663636	0,6637	18
F4	4,472727	0,868182	0,8682	6
F5	3,854545	0,713636	0,7137	16
F6	3,690909	0,672727	0,6728	17
F7	2,8	0,45	0,45	21
F8	2,963636	0,490909	0,491	20
F9	4,345455	0,836364	0,8364	10
F10	4,418182	0,854545	0,8546	9
F11	4,309091	0,827273	0,8273	11
F12	3,345455	0,586364	0,5864	19

3.6. Measure Improvement Ratio

At this stage, it is necessary to determine the goals of each feature. There are two ways of determining goals: according to consumer desires, company policies, and the company's capabilities. In this study, the parameters of nominal goals consider the results of benchmarking the performance value between products and competitors subjectively (dummy). The improvement ratio (IRt) is calculated by comparing goals (G) with current performance (CP). The results of the IRt using goals by consumer desires can be seen in Table 8.

In the following process, namely determining the point of sale, determining the scale of consumer interest, determining a list of technical needs, analyzing the relationship between customer needs and technical needs, calculating the technical needs score, determining the correlation between technical requirements and compiling all calculation and analysis data into the House of Quality (HoQ) matrix, are needed.

Table 8. Example of IRt measure

Feature	CP	G	IRt
F1	0,046	0,05	1,094092

F2	0,032	0,05	1,567398
F3	0,036	0,001	0,02809
F4	0,095	0,001	0,010549
F5	0,016	0,001	0,062893
F6	0,079	0,001	0,012626
F7	0,095	0,001	0,010526
F8	0,02	0,001	0,050505
F9	0,145	0,05	0,345781
F10	0,026	0,05	1,937984
F11	0,036	0,05	1,40056
F12	0,003	0,001	0,344828

By the criteria obtained, namely the highest performance value, the average result of 10-fold cross-validation, shows 0.850463, and the ROC result shows 0.858485 and 0.866732. Then, Document B, with a performance of 71, is the model to be selected. Previously, several experiments have been carried out to optimize the performance of existing datasets.

The distribution in each dataset should be balanced and contain all cases [16]. For example, a binary classification dataset should contain 50% positive and 50% negative cases. The number of data was 1673 in the review; previously, it has experienced a gradual addition until finally getting this value. There are three documents, namely A, B, and C, with different numbers of datasets, and each of them has the same class proportions. The three of them then carried out the same process to determine the trend of model performance to the amount of data. According to the results, the average optimal performance is in Document B. It decreases in document C. Then, no additional data is added, and it moves to the next stage of optimizing the model's performance, namely the pre-processing stage. In pre-processing, in addition to adding stop-words and normalization that can increase performance values, there is a process that can reduce noise in the classification process to improve performance, namely the stemming process. However, popular libraries such as NLTK and spacy are not yet available in Indonesian, this causes some words to be inappropriate when changed to their root words. The vectorization process with TFIDF and BoW does not consider compound words and word phrases.

Based on information theory or entropy, feature selection is carried out on features that are considered to have no high information value. In other words, the feature that cannot differentiate one class from another is uninformative. However, the use of popular feature selection, namely MI and TFIDF L1, does not match the dataset; this is because some features are considered undesirable in the MI and TFIDF L1 processes but are actually needed in the classification process.

Many data mining or machine learning techniques are developed with the assumption of linearity, so that the resulting algorithm is limited to linear cases [17]. SVM can work on non-linear data by using a kernel approach to the initial data set feature. A kernel function used to map the initial dimension (lower dimension) of a data set to a new dimension (a relatively higher dimension). Which kernel function should be used for dot product substitution in feature space is highly dependent on the data because this kernel function will determine the new feature in which the hyperplane will be searched. In the first, second, and third experiments the kernel used is linear because when the RBF kernel is used with the default parameters used, performance will decrease. However, with a combination of parameters in the hyperparameter, the RBF kernel has the highest performance value. The model is a multi-label classification, a soft classification that is able to classify a term into several classes, for example term x has a 70% probability as class A and 30% as class B. In the data it is found that some of these words are not specific in one class. so that it can be found in more than one class and this affects the algorithm in predicting data.

4. Conclusion

The product development process is carried out with the initial step, namely design planning, in terms of physical and quality. Sentiment analysis will be used to find the class group of each data. The first process is manual labeling, which provides an initial identity to the supervised learning algorithm. The QFD stage will use a negative class as a feature of consumer needs. Word association is carried out with the collocations function to obtain information from predicted data. The t-test was chosen because it has the most significant true positive value of the five collocation methods. As a comparison, a data search process was carried out on a competitor's product, namely the Converse Chuck Taylor 70s. Competitor product data goes through the same process as Ventela Public Low's product review data. The first process in QFD is done by matching data and grouping association data with terms with the same information. The second process is calculating product performance; in this integrated model, the product performance value will be calculated from its association value. The weakness in this calculation is if there is data with terms with the same information but very different association values due to the appearance of different terms. This will have an effect because the value is calculated on average, so if there are equal terms with different values, that can significantly reduce the value. To obtain a balanced value, weighting is carried out for each feature based on the amount of constituent data. Furthermore, when making features or attributes or criteria that represent the constituent data, starting in this process, the role of stakeholders is needed to analyze the data results. In the next stage of IR calculation, because the importance of a feature needs to be assessed directly, the association value cannot be used for this process. In the following process, namely the improvement ratio, determining the point of sale, determining the scale of consumer interest, determining the list of technical needs, analyzing the relationship between customer needs and technical needs, calculating the score of technical needs, determining the correlation between technical requirements and compiling all calculation and analysis data into a matrix HoQ.

From the simulation results, the researcher uses dummy data, showing that the integration between review data analysis using machine learning and QFD produces appropriate information by the provisions of the QFD method and can support the product development process in terms of the amount of data, various data topics, and reduce subjectivity at this stage—determination of VoC and product performance and competitors. For further research, a text processing algorithm in Indonesian can be developed to capture words according to the characteristics of Indonesian, using other classification algorithms such as linear regression, decision tree, and logistic regression. This research will be integrated with other methods to optimize and reduce subjectivity in QFD such as fuzzy logic, artificial neural networks, the Taguchi method, and other methods and developing an approach in the form of a calculation formulation that involves the results of benchmarking values to apply them to the priority ranking decision criteria that need to be improved.

References

[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] IMARKETOLOGY, "Apa yang Diungkapkan Data Tentang Pasar Online Indonesia dan Global?," academy.getcraft.com, 21 April 2020. [Online]. Available: https://academy.getcraft.com/id/blog/apa-yang-diungkapkan-data-tentang-pasar-online-indonesia-dan-global. [Accessed 21 April 2020].

[3] A. Farki, I. Baihaqi and B. M. Wibawa, 2016, "Pengaruh Online Customer Review dan Rating Terhadap Kepercayaan dan Minat Pembelian pada Online Marketplace di Indonesia," Jurnal Teknik ITS, vol. 5, no. 2, pp. A614-1619.

[4] H., 2016, "Big Data Analisis: Sebuah Peluang dan Tantangan Lulusan Informatika Dalam Menghadapi Pasar Global di Indonesia Timur," in Proceedings Seminar Nasional Riset Ilmu Komputer (SNRIK), Makassar.

[5] Y. A. Krisdamarjati, "Ventela dan Tantangan Popularitas Sepatu Lokal di Indonesia," kompas.id, 29 Februari 2020. [Online]. Available:

https://kompas.id/baca/riset/2020/02/29/ventela-dan-tantangan-popularitas-sepatu-lokal-di-indonesia/. [Accessed 06 Agustus 2020]..

[6] C. Resmi, "Ventela, Sepatu Lokal Bandung yang Mulai Hits di Kalangan Milenial," urbanasia.com, 25 November 2019. [Online]. Available: https://www.urbanasia.com/ventela-sepatu-lokal-bandung-yang-mulai-hits-di-kalangan-milenial-U5995. [Accessed 06 Agustus 2020].

[7] indoworx.com, "Inilah 7 Produk Gagal Dari Perusahaan Terkenal," indoworx.com, 14 Mei 2018. [Online]. Available: https://www.indoworx.com/produk-gagal-dari-perusahaan-terkenal/. [Accessed 21 April 2020].

[8] F. Sebastiani, 2002, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-49.

[9] N D. M. Powers, 2007, "Evaluation: From Precision, Recall, and F-Factor," Technical Report SIE-07-001, pp. 1-22.

[10] W. Arini, 2003, Manajemen Kualitas, Jakarta: Ghalia Indonesia.

[11] D. L. Goetsch and S. B. Davis, 2000, Quality Management: Introduction to Total Quality Management for Production Processing and Services, New Jersey: Pretince Hall.

[12] V. Gasperz, 1988, Manajemen Kualitas Dalam Industri Jasa, Jakarta: Gramedia Pustaka Umum.

[13] F. Gorunescu, 2011, Data Mining Concept, Models, and Techniques, Berlin: Springer.

[14] S. Evert and B. Krenn, 2001, "Methods for Qualitative Evaluation of Lexical Association Measures," in Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Toulouse.

[15] S. Evert, 2005, The Statistics of Word Cooccurrences Word Pairs and Collocation, Stuttgart: University of Stuttgart.

[16] E. A. Cudney and C. C. Elrod, 2011, "Quality Function Deployment in Continuous Improvement," intechopen.com, pp. 3-78.

[17] J. W. G. Putra, 2020, Pengenalan Konsep Pembelajaran Mesin dan Deep Learning Edisi 1.3, Tokyo: self-published-work.

[18] B. Santosa, 2007, Data Mining Teknik Pemanfaatn Data Untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu.