Advances in Economics, Business and Management Research, volume 202
3rd International Conference on Business and Management of Technology (ICONBMT 2021)

Sentiment Analysis Based on Quality Aspects in Effort to Improve Quality of Indhome Product and Services PT Telkom Indonesia Tbk

Reza Hermansyah[1,*], Riyanarto Sarno[1,*]

[1] Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[*]Corresponding author. Email: rezahrmnsyh@gmail.com, riyanarto@if.its.ac.id

ABSTRACT

The fierce competition for product quality in the telecommunications industry makes every company compete to provide products that are acceptable and liked by the public, including PT Telkom Indonesia Tbk with one of its superior products, Indihome. To find out the advantages of a product or service, it is necessary to evaluate the opinion of the online review. One method that is often used to analyze online reviews is sentiment analysis. This study conducted a sentiment analysis on the product or service Indihome PT. Telkom Indonesia Tbk is based on a review written in tweet on Twitter. First, Each tweet related to Indihome is collected into data and then categorized based on eight aspects of the dimension of quality (Performance, Reliability, Features, Fit & Finish, Serviceability, Durability, Conformance dan Aesthetics) with several stages. First, the keyword term aspect of the product quality dimensions are expanded with the help of the Term Frequency Inverse Cluster Frequency method or better known as TF-ICF. Then, Lantent Dirichelet Allocation (LDA) to find hidden topics from tweets and calculate Sematic Similarity to categorize each hidden topic. Furthermore, the comparison of the combination of classification algorithm types: SentiWordNet, SentiCircle, SVM and Random Forest was tested using an evaluation factor, this was done to get a sentiment model with the best performance. The results show that the categorization of aspects using LDA+Semantic Similarity model and combining it with TF-ICF 100% to expand the terms gave 70.1% results. The sentiment classification shows that the combined SentiWordNet+SVM model assisted by SentiCircle produces a performance of 96.3%. Finally, the researchers used these two methods to obtain an evaluation based on eight aspects of the dimension of quality towards improving the quality of Indihome's products and services. The researcher found that the ease of service aspect (Serviceability) had a review with the highest negative sentiment reaching 20.82% compared to other dimensions (Performance : 12.86%, Reliability: 2.53%, Features: 13.28%, Fit & Finish : 0.99%, Durability 4.44%, Suitability : 2.29% and Aesthetics : 5.86%).

Keywords: *Online Review, Sentiment Analysis,, Dimension of Quality, Indihome*

1. INTRODUCTION

The tight competition in the telecommunications industry in Indonesia requires every company to have an advantage over the products or services offered to remain competitive [1]. The choice of products and services is increasingly varied, marked by the number of companies engaged in the telecommunications industry. One of the efforts that companies can make to outperform their competitors is to analyze customer opinions on the products and services they offer.

PT Telkom Indonesia Tbk or commonly called Telkom is one of the state-owned companies engaged in telecommunications services in Indonesia. Telkom's current business focus is digital connectivity with Indihome as its main product [2]. The competitive market climate requires Telkom to measure the opinions of Indihome customers to retain existing customers and attract new potential customers.

Online reviews are a benchmark for the value of a brand's opinion on products, services, or experiences, it becomes important for businesses to maintain their quality in the market [3], [4]. One method that is often used to analyze online reviews is sentiment analysis [5] [6]. Due to its ability to gain rich insight into the details and reasons for unclear market trends, sentiment analysis techniques are widely used

Social media is an example of the application of popular sentiment analysis to find out opinions through online reviews, this is supported by the rapid development of internet use among the public. One of the popular social media to express opinions or opinions today is Twitter. Based on the results of a survey on the most popular social media in Indonesia in January 2021, Twitter was ranked fifth with 63.6% [7]. Through Twitter, users can send and read text-based messages known as tweets, in addition to text types of messages that can be posted via tweets in the form of photos, videos, links.

This study conducted a sentiment analysis of online reviews through Twitter-based on aspects of products and services. The selected aspects are based on eight aspects of the quality dimension (Performance, Features, Reliability, Suitability, Durability, Ease of Service, Aesthetics & Perceived Quality). To categorize these aspects, this study uses the help of the LDA, Sematic Similarity and TF-ICF methods. LDA is a method used as topic modeling in detecting hidden topics from a document [8]. The term list before entering the semantic similarity stage has been determined beforehand and expanded using TF-ICF.

Then, the combination of SentiwordNet, SentiCircle, SVM and Random Forest classification methods is compared to determine which model produces the best performance in the sentiment classification process. The result of this process is a positive or a negative opinion. Furthermore, the calculation of Net Brand Reputation (NBR) is carried out to measure customer loyalty to the product category owned by a brand on social media.

Based on this, the expected final result in this research is an analysis of the suitability of the quality dimension aspects of improving the quality of Indihome products and services owned by Telkom. In addition, the acquisition of sentiment analysis is also expected to be a benchmark for measuring online reviews on company social media as a complement to the Net Promoter Score (NPS) method which has been routinely carried out and can be implemented by the Customer Marketing and Customer Care units.

2. LITERATURE REVIEW

2.1 Previous Research

In a previous study entitled "Twitter Sentiment to Analyse Net Brand Reputation of Mobile Phone Providers", sentiment analysis technique was used to classify the reputation of a telecommunications brand [9]. This study calculates the reputation of PT Axiata Tbk (XL) products and compares them with products owned by Telkomsel and Indosat. Furthermore, by comparing three classification algorithms Naïve Bayes, Decision Tree, and Support Vector Machine. Support Vector Machine (SVM) is used to calculate Net Brand Reputation (NBR) because it provides better performance than other models. NBR calculations are carried out on 3G, 4G, Voice, SMS and Internet (data) products. The results obtained from the study stated that XL Axiata obtained a higher average reputation value based on a comparison of five types of products by 32.3% compared to Telkomsel which only obtained 19.0% and Indosat 10.9%.

In previous studies sentiment analysis technique used aspect based in classifying sentiment in the hotel industry through Twitter social media [10]. In an effort to find the best aspect categorization performance, this research is carried out by combining several methods into several implementations. The highest aspect categorization model reached 85% using LDA and sematic similarity then using TF-ICF with a rate of 100%.

2.2 The Eight Aspects of the Dimensions of Quality

Quality is an important factor in determining a product. With the quality of the company can measure the level of good or bad products they produce. In 1987 David A. Gavin introduced eight aspects of quality dimensions that aim to solve the ambiguity and subjectivity that arise in measuring product quality, the eight dimensions of quality include: Performance, Durability, Features, Conformance, Serviceability, Reliability, Perceived Quality and Aesthetics [11].

Table 1 Definitions of Aspect Based on Keyword Term

Aspect of Quality Dimension	Term Variables
Performance	‘Lemot’, ‘Lambat’, ‘Lama’, ‘Cepat’, ‘Lancar’, ‘Normal’, ‘Hilang’, ‘Unduh’, ‘Unggah’
Features	‘Wifi’, ‘useetv’, ‘game’, ‘telkomsel’, ‘Netflix’, ‘poin’, ‘telepon’, ‘zoom’, ‘instagram’, ‘drakor’
Reliability	‘Putus’, ‘Patah’, ‘Kedip’, ‘Stabil’, ‘Ganggu’
Conformance	‘Tagihan’, ‘Mahal’, ‘Murah’, ‘Bohong’, ‘Jujur’, ‘Aman’, ‘Bapuk’
Durability	‘Rusak’, ‘Awet’, ‘Tahan’, ‘Usia’, ‘Waktu’
Serviceability	‘Professional’, ‘Ramah’, ‘Respon’, ‘Komunikasi’, ‘Solusi’, ‘Prilaku’
Aesthetics	‘Kotor’, ‘Bersih’, ‘Rapih’, ‘Indah’, ‘Buruk’
Perceived Quality	‘Puas’, ‘Senang’, ‘Males’, ‘Emosi’, ‘Bosan’, ‘Bahagia’, ‘Rekomendasi’, ‘Juara’, ‘Henti’

2.3 Pre-processing

Preprocessing is a data mining technique used to convert raw data into a useful and efficient format [12]. Table 2.2 can be seen the types and descriptions of the preprocessing process used in this study.

Table 2 Description Each Type of Preprocessing

Preprocessing	Explanation
Convert to Lowercase	Convert to Lowercase Converts each word to lowercase
Spelling Correction	Identifying and correcting spelling errors in words in a document
Punctuation Removal	Process to remove all unwanted punctuation marks like apostrophes, brackets, colons, commas, dashes, etc.
Stopword Removal	Delete a set of words that are commonly used in any type of language.
Tokenization	Process of converting to token before converting it to vector
Stemming	Method for removing suffixes from words and bringing them to the root

2.4 Latent Dirichlet Allocation (LDA)

LDA is a topic modeling method by detecting hidden topics in the document and the proportion of their occurrence [13]. LDA can determine the number of topics from a corpus of documents and the distribution of words in each topic. Following are the steps in the LDA process :

1. Select the sample document length N from the Poission distribution (ξ)
2. Select distribution θd on topic from dirichlet distribution (α)
3. N represents each word i.e. w_n :
4. Select topic z_n from Multinominal (θ)
5. Choose the word w_n from P(w_n | z_n ,β), the multinominal probability on the topic

$$p(w,z,\theta,\varphi;\alpha,\beta) = \prod_{j=1}^{M} p(\theta_j|\alpha) \prod_{i=1}^{K} p(\theta_i|\beta) \prod_{t=1}^{N} p(z_{j,t}|\theta_j) p(w_{j,t}|\varphi_i) \quad (1)$$

witch :

β = dirichlet parameter on word distribution to topic

φ = distribution of words against topics in the corpus

K = topic group

W = word

N = word group

M = document set

z = topic index assignment

θ = document

α = dirichlet parameter on the distribution of topics to the document.

2.5 Sematic Similarity

To give a numerical similarity score to words to represent the distance of their sentiments, the Semantic Similarity method can be used. Semantic relationship assumes that two objects are semantically related if they have some type of semantic relationship [14]. Similarity distance measures the proximity between word 1 ($\boldsymbol{w_i}$) and word 2 ($\boldsymbol{w_j}$) can be seen in Equation (2). The similarity value ranges from 0 to 1.

$$Similarity\,(w_i, w_j) = \frac{\sum_{m-1}^{K} w_i^m w_j^m}{\sqrt{\sum_{m-1}^{K} (w_i^m)^2}\sqrt{\sum_{m-1}^{K} (w_j^m)^2}} \quad (2)$$

Equation 2 was used to measure the similarity between ($\boldsymbol{w_i}$) term 1 and ($\boldsymbol{w_j}$) term 2. Where $\sum_{m-1}^{k}$ is number of iterations from m to K word.

2.6 TF-ICF

The function of the TF-ICF is to weight each term against the information possessed by a document or text in a cluster. Information is obtained by calculating the frequency value of the term against the cluster using the equation (3) and (4).

$$TF - ICF_i = tf_{ij}\ x\ ICF_I \quad (3)$$

$$ICF_i = 1 + log\frac{c}{cf_i} \quad (4)$$

witch :

ICF_i = inverse class frequency for each term/word

c = the number of classes in the training data

cf_i = number of classes that have the word

2.7 SentiWordNet

SentiWordNet is a lexical database and term trend scores to support the classification of positive or negative sentiments in the opinion mining process [15]. Each synset in WordNet has a numerical score in the range [0,0;1,0]. SentiWordNet which is a semi-supervised classifier learning method was developed from WordNet to meet research needs [16].

The results from SentiWordNet, will get the polarity weight for each word where the weight will be used to label the word in the next stage. In the use of SentiWordNet there are four POS Tags are used so it is necessary to change the POS Tag format from Indonesian to SentiWordNet

2.8 SentiCircle

SentiCircle [17] aims to study the sentiment orientation of words from their contextual semantics. This method assumes that sentiment is a term that can change depending on the context in which the term is used [18]. The senticircle method represents the sentiment model using a polar coordinate system [19] which can be seen in Figure 1.

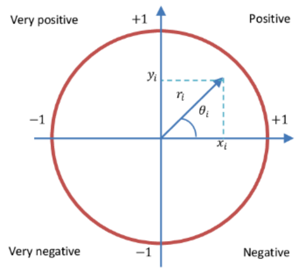

Figure 1 SentiCircle Representation

Sentiment polarity value is represented by the y-axis and the strength of sentiment is represented by the x-axis. First, the Term Degree of Correlation (TDOC) needs to be determined to calculate the polarity value of a word using the equation (5). Next, determine the radius for each context term using Equation (6). Then determine the angle of each context term with Equation (7). The last step is to determine the x and y positions of each context using Equations (8) and (9). Finally, to get the SentiMedian which is the final result of the polarity value using Equation (10).

$$TDOC(m, c_i)_d = f(c_i, m) \times log \frac{N}{N_{ci}} \quad (5)$$

$$r_i = TDOC(m, c_i)_d \quad (6)$$

$$\theta_i = Prior\_sent(c_i) \times \pi \quad (7)$$

$$x_i = r_i cos\theta_i \quad (8)$$

$$y_i = r_i sin\theta_i \quad (9)$$

$$g = \arg\min_{g \in R^2} \sum_{i=1}^{n} \|p_i - g\|_2 \quad (10)$$

witch :

d = document

m = opinion word

c_i = opinion on the context of a word

N = the number of opinion words (m) contained in the document

N_{c_i} = the number of opinions on the context of a word (c_i) in the document

$f(c_i, m)$ = frequency of joint between m and c_i

r_i = radius

P_sent = polaritas value

θ_i = c_i degree

x_i and y_i= c_i position in x and y axis

p_i = c_i position

g = position of m (x_m and y_m)

2.9 Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the methods in supervised learning. SVM is used to find the best hyperplane by maximizing the distance between classes [20]. In SVM to separate between classes using a function called hyperplane [21]. The function used for classification between classes in 2 dimensions is called a line and in 3 dimensions is called a similar plane, while hyperplane is used for classification in a higher

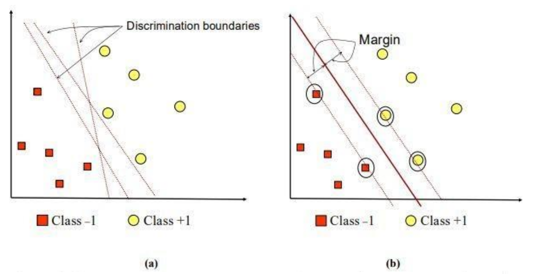

Figure 2 SVM Hyperplane

dimensional class space.Image source [22]

2.10 Random Forest

Random Forest is a supervised classification method that is part of the decision tree method [23]. The decision tree can classify a data sample whose class is not yet known into existing classes. Flowcharts such as trees with root nodes, inner leaf and leaf nodes are used to solve problems and make decisions.

Random Forest is a combination of good tree results that are combined into one model. Random Forest depends on a random vector value with the same distribution in all trees where each decision tree has a maximum depth.

3. METHOD

The method proposed in this study can be seen in Figure 3.

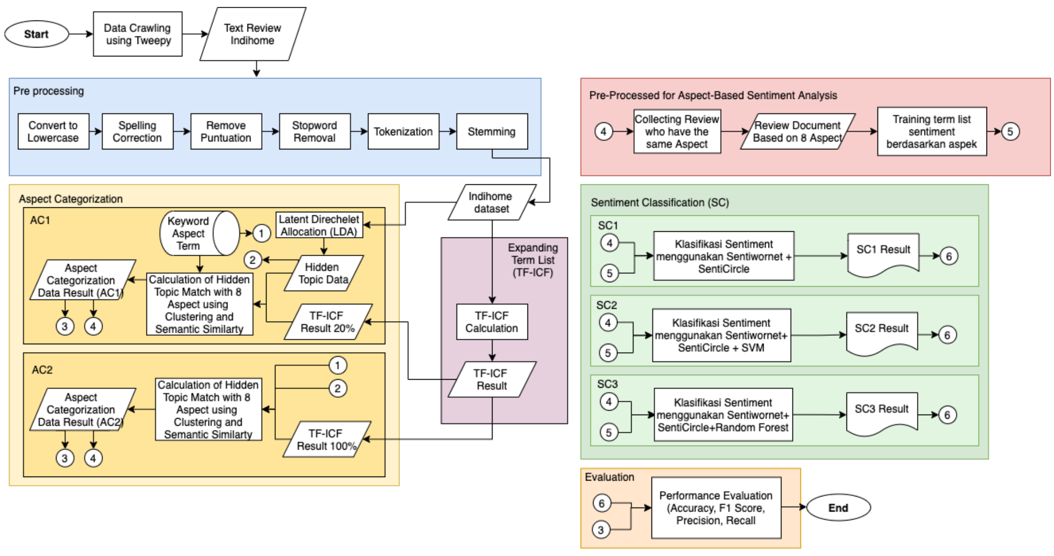

Figure 3 Flowchart of Research method

Firstly, Tweet data in the form of the keyword "Indohome" will be crawled using the API owned by Twitter using Tweepy. After the data is obtained and saved in CSV file format. Data obtained is then through the preprocessing process, starting from removing punctuation marks, trimming spaces between words to removing punctuation marks. The results of the pre-processing stage will process each review into a term list. Furthermore, the dataset is divided into 80% training data and 20% into testing data.

As an effort to expand the term list and improve performance when calculating Semantic Similarity, the training data will be processed using TF-ICF. With this method, important terms from the new quality dimension aspects will be added, so as to optimize the aspect categorization process, especially in the calculation of hidden topics and semantic similarity.

Afterward, the LDA method is used to find hidden topics followed by the Sematic Similarity equation to calculate topic categorization based on the aspect of quality dimensions. In this process, the data will be categorized into eight aspects of the dimensions of quality. Then data that has the same glossary of terms are collected based on the results of the aspect categorization and trained to get sentiments with the same aspect category. The result of the term list is used as input to find sentiment classification based on dimensional aspects.

Sentiment classification is done by a combination of SentiWordNet, SVM and Random Forest algorithms. The results of this process are positive and negative labels per aspect of the quality dimension of the Indihome review. As an effort to increase the accuracy of sentiment on aspects of the SentiCircle method, the method is used to determine the new polarity value of each sentiment classification model. To get the best method performance, each process on sentiment categorization and classification will go through several tests and will be evaluated using accuracy, precision, recall and F1-measure. Aspect categorization is divided into AC1 and AC2 models and for aspect classification is divided into SC1, SC2 and SC3 models.

4. RESULT AND ANALYSIS

4.1 Crawling Data

The data used in this study are consumer opinions written online in the form of tweets from Twitter. The tweets have taken by keyword "Indihome". An example of the process and results of the data crawling process can be seen in Table 3.

Table 3 Crawling Data Result

ID Review	Tweet Id	Text
1	1376552614981115907	Kau enak Internet mu cepat. Lah aku server tokyo +indihome : aib #PS4share #indihome
2	1361279478081620000	@HappyLiey Hari senin + Internet lancar dan stabil = Surga dunia #IndiHome

4.2 Pre-processing

Table 4 is an illustration of the results of data processed using pre-processing steps

Table 4 Illustration of Pre-processing Result

ID Review	Hasil Preprocesing
1	"tokyo", " internet", " enak", " aib", " cepat" "ps4share"
2	"lancar", "stabil", "senin", "surga", "internet", "dunia"

4.3 Expanding Term List

In this process, it is used to help the limitations of the previously defined terms [24]. The results obtained in this process will be used as a calculation process in the Semantic Similarity process. The process at this stage begins by calculating the TF-ICF for each cluster based on the term list which will be expanded in Table 5.

It can be seen in Table 5 that the TF-ICF value will be greater if the ICF value is not equal to zero. This proves that the larger a word appears in a limited cluster, it is directly proportional to the possibility that the word can represent a cluster.

Table 5 Calculation of Expanding Term List (TF-ICF)

Id_term	Term	TF-ICF *Performance*	TF-ICF *Reliability*	TF-ICF *Features*	TF-ICF *Fit & Finish*	TF-ICF *Serviceability*	TF-ICF *Durability*	TF-ICF *Conformance*	TF-ICF *Aesthetics*
1	internet	0	0.192	0.256	0.192	0.256	0.064	0.064	0
2	enak	0	0.333	0	0	0	0.333	0	0
3	cepat	0	0.226	0.226	0.226	0.226	0.452	0	0.226

4.4 Aspect Categorization

The aspect categorization process can be carried out after the pre-processing and expanding term list stages are completed. Eight aspects of the quality dimension will be looked for in each tweet at this stage. The categorization of aspects in this study will compare the test results between Aspect Categorization 1 (AC1) and Aspect Categorization 2 (AC2). The second test uses the LDA to generate hidden topics from the data and is calculated using the Semantic Similarity method. However, the difference between the two Tests is the vocabulary expansion aid rate used. AC1 uses 20% TF-ICF and AC2 uses 100% TF-ICF extension.

4.4.1. Latent Dirichlet Allocation (LDA)

In its implementation, LDA uses the term frequency to build a vector of each word token to find the hidden topic of a document. Python's Gensim library helps in constructing the LDA model. The following tests on the research are presented in Table 6

Table 6 LDA Hidden Topic Result

ID Review	*Hidden Topic*
1	0.216*"internet"+ 0.110*"cepat" + 0.012*"tokyo" + 0.012*"aib" + 0.012*"ps4share" + 0.011*"enak"
2	0.009*"lancar"+ 0.261*"stabil"+ 0.013*"senin" + 0.016*"surga"+0.239*"internet"+ 0.013*"dunia"

4.4.2. Semantic Similarity

After obtaining hidden topics from LDA for each tweet data in Table 6 and adding data from the TF-ICF process in Table 5, the results in the form of a list of words are processed in the calculation of semantic similarity. Hidden topics that are very similar are most likely to represent a predetermined aspect, the value scale is given from 0 to 1. A value of 0 indicates that the topic does not represent an aspect and a value of 1 otherwise.

Table 7 LDA + Semantic Similarity + TF-ICF Result

LDA + *Sematic Similarity* (TF-ICF 20%)										
ID Review	**A0** ***Performance***	**A1** ***Reliability***	**A2** ***Features***	**A3** ***Fit & Finish***	**A4** ***Serviceability***	**A5** ***Durability***	**A6** ***Conformance***	**A7** ***Aesthetics***	**Label**	**Label Anotator**
1	**0.117**	0.093	0.085	0.000	0.142	0.142	0.062	0.111	*Performance*	*Reliability*
2	0.206	0.093	0.085	0.140	**0.260**	0.260	0.111	0.199	*Serviceability*	*Reliability*
LDA + *Sematic Similarity* (TF-ICF 100%)										
ID Review	**A0** ***Performance***	**A1** ***Reliability***	**A2** ***Features***	**A3** ***Fit & Finish***	**A4** ***Serviceability***	**A5** ***Durability***	**A6** ***Conformance***	**A7** ***Aesthetics***	**Label**	**Label Anotator**
1	0.117	**0.253**	0.245	0.000	0.142	0.142	0.062	0.111	*Reliability*	*Reliability*
2	0.206	0.200	0.200	0.200	**0.260**	0.260	0.111	0.199	*Serviceability*	*(Reliability*

The performance of the categorization of aspects will be calculated with the evaluation parameter in the form of an F1-measure. The following are the test results on the performance of each aspect categorization model:

Table 8 Aspect Categorization Model Performance

Performance Aspect Analysis		
Model Approach	**Performance Calculation**	**F1-Measure**
AC1	LDA + *Semantic Similarity* + TF-ICF 20%	0.693
AC2	LDA + *Semantic Similarity* + TF-ICF 100%	**0.701**

Table 6 shows that the addition of the term list to the TF-ICF with a rate of 100% which can find terms that represent every aspect of the document and make the semantic similarity process better.

4.5 Sentiment Classification

The classification method used is a combination of the SentiWordNet, SVM and Random Forest algorithms. The results of each sentiment classification model will then be re-determined using SentiCircle polarity in the hope of increasing its accuracy.

The SC1 test uses the SentiWordNet + SentiCicle classification, SC2 is a combination of SC1+SVM models, and SC3 is a combination of SC1+Random Forest models. In this study, the classification used is binary classification, where the classification results are divided into two types of class labels, namely positive and negative. The following are the results of the F1-measure test on the performance of each sentiment classification model:

Table 9 Sentiment Classification Performance

Sentiment Analysis Performance		
Model Approach	**Performance Calculation**	**F1-Measure**
SC1	SentiWordNet + SentiCircle	0.643
SC2	SentiWordNet + SentiCircle + SVM	**0.963**
SC3	SentiWordNet + SentiCircle + Random Forest	**0.963**

Based on the F1-measure value obtained in Table 9, further analysis needs to be done to select a sentiment classification model between SC2 and SC3 models which has the same F1-measure value of 0.963 using Stratified K-Fold = 5.

Table 10 Sentiment Classification Performance (SC2 vs SC3)

Sentiment Analysis Performance		
Performance Calculation	**Sentiment Classification**	
	SC2	**SC3**
Precision	**93.9%**	93.8%
Recall	**93.7%**	93.6%
F1-Measure	**93.6%**	93.6%

Table 10 shows that the combination of SentiWordNet + SentiCircle + SVM method has better accuracy, precision and recall performance compared to SC3 which uses SentiWordNet + SentiCircle +

Random Forest with an average performance difference of 0.1%.

After knowing the best performance from each test model, the next step is to conduct a sentiment analysis based on the aspect of quality dimensions on Indihome. With this, the aspect categorization model in the 2nd experiment (AC2) and sentiment classification in the 2nd experiment (SC2) coupled with the improvement in SentiCircle accuracy is used to perform sentiment analysis. The following are the results of sentiment evaluation based on aspects of the quality dimensions of Indihome's products and services :

Table 11 Sentiment Evaluaton Result on Indihome

Sentiment Analysis Results		
Aspect	**Sentiment**	***Evaluation* Results**
Performance	Positive	2.44%
	Negative	12.86%
Reliability	Positive	2.44%
	Negative	2.53%
Features	Positive	5.29%
	Negative	13.28%
Fit and Finish	Positive	0.10%
	Negative	0.99%
Serviceability	Positive	18.30%
	Negative	20.82%
Durability	Positive	2.75%
	Negative	4.44%
Conformance	Positive	1.52%
	Negative	2.29%
Aesthetics	Positive	4.11%
	Negative	5.86%
Total *Percentage*		100%

Table 12 Examples of the Influence of Aspects on a Change in a Sentiment

Review	Aspect	Sentiment
Indihome Kecepatan tinggi yg tak di ragukan	*Performance*	**Positive**
Internet sampah guys... Bayar mahal, dpt kualitas murahan. Gangguan Indihome tinggi sekali	*Serviceability*	**Negative**

From Table 12 in the review example, the word "high" in a sentence has two different meanings. The first word high means satisfied because the internet is fast. In the next review, the "high" then has a different aspect. The first word is high on the performance aspect because the customer's internet speed is very good and the customer is satisfied, while the second "high" refers to the Serviceability aspect and has a bad meaning because Indihome's disturbances often occur.

5. CONCLUSION

1. Based on several testing methods that have been implemented and analysis of the test results obtained, several conclusions can be drawn. First, the addition of TF-ICF which is tasked with expanding the terms that represent every aspect of the document and making the aspect categorization process better. In the experiment, AC1 applied 20% of the total TF-ICF found and AC2 applied 100% of the TF-ICF rate. In the performance evaluation, it was found that the AC2 trial was better than the AC1 trial with an application performance of 70.1% for AC2 and 69.3% for AC1. So, it can be concluded that applying an additional rate of TF-ICF can show the best results because it can increase performance by 0.8%. The accuracy for Sentiment Classification with the best performance using the SentiWordNet + SentiCircle + SVM increased by 93.7%.

2. The result shows that the negative sentiment towards the aspect of Serviceability dimension of 20.82% indicates that it shows that customers really care about competence, speed, accuracy, and convenience in providing services for Indihome improvements which still need to be improved. Steps that can be taken by Indihome are to improve the quality of its technicians, customer relations staff when dealing with customer complaints and disturbances by holding training programs and customer retention.

3. Negative sentiment on the aspect of the feature dimension which reached 13.28% indicates that the additional characteristics that complement the benefits of Indihome do not provide significant added value for customers. The Indihome team needs to focus on developing product features or additional features offered to customers, so that customers don't feel that the features that Indihome has are just a gimmick.

4. Negative sentiment towards aspects of the performance dimension with 12.86% shows that the functional aspects and main characteristics of Indihome are still relatively bad. The visible solution for Telkom's management is to re-audit the quality

of people, processes and technology on an end-to-end basis.

5. The results also show that the aspect of each review affects the change in sentiment (positive or negative sentiment).

AUTHORS' CONTRIBUTIONS

Reza Hermansyah : study concept or design, data collection, write the paper. Riyanarto Sarno: supervise, and review the manuscript.

ACKNOWLEDGMENTS

This research was funded by the Indonesian Ministry of Education and Culture under Riset Inovatif-Produktif (RISPRO) Invitation Program managed by Lembaga Pengelola Dana Pendidikan (LPDP) and under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program managed by Institut Teknologi Sepuluh Nopember (ITS).

REFERENCES

[1] E. Sutarsih, V. C. Wulandari, R. Utari, N. A. Rozama, dan Kusumatrisna, *Statistik Telekomunikasi Indonesia 2019*. Badan Pusat Statistik, 2020. [Daring]. Tersedia pada: https://www.bps.go.id/publication/2020/12/02/be999725b7aeee62d84c6660/statistik-telekomunikasi-indonesia-2019.html

[2] PT Telkom Indonesia Tbk, "Profil dan Riwayat Singkat," 2020. https://telkom.co.id/sites/about-telkom/id_ID/page/profil-dan-riwayat-singkat-22

[3] A. Ghazian, M. H. Hossaini, dan H. Farsijani, "The Effect of Customer Relationship Management and its Significant Relationship by Customers' Reactions in LG Company," *Procedia Economics and Finance*, vol. 36, hlm. 42–50, 2016, doi: 10.1016/S2212-5671(16)30014-4.

[4] I. Buil, E. Martínez, dan L. de Chernatony, "The influence of brand equity on consumer responses," *Journal of Consumer Marketing*, vol. 30, no. 1, hlm. 62–74, Jan 2013, doi: 10.1108/07363761311290849.

[5] R. A. Laksono, K. R. Sungkono, R. Sarno, dan C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," dalam *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Surabaya, Indonesia, Jul 2019, hlm. 49–54. doi: 10.1109/ICTS.2019.8850982.

[6] B. Pang dan L. Lee, "Opinion mining and sentiment analysis," hlm. 94.

[7] S. Kemp, "Digital 2021 Indonesia," *Digital 2021 Indonesia*, Feb 11, 2021. https://datareportal.com/reports/digital-2021-indonesia

[8] V. Bonta, N. Kumaresh, dan N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," hlm. 6, 2019.

[9] N. A. Vidya, M. I. Fanany, dan I. Budi, "Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers," *Procedia Computer Science*, vol. 72, hlm. 519–526, 2015, doi: 10.1016/j.procs.2015.12.159.

[10] R. Priyantina dan R. Sarno, "Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM," *IJIES*, vol. 12, no. 4, hlm. 142–155, Agu 2019, doi: 10.22266/ijies2019.0831.14.

[11] David. A. Garvin, "Competing on the Eight Dimensions of Quality," *Harvard Business Review*, vol. 65, no. 6, Des 1987.

[12] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*. Elsevier.

[13] Q. Chen, L. Yao, dan J. Yang, "Short text classification based on LDA topic model," dalam *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, Shanghai, China, Jul 2016, hlm. 749–753. doi: 10.1109/ICALIP.2016.7846525.

[14] O. Araque, G. Zhu, dan C. A. Iglesias, "A semantic similarity-based perspective of affect lexicons for sentiment analysis," *Knowledge-Based Systems*, vol. 165, hlm. 346–359, Feb 2019, doi: 10.1016/j.knosys.2018.12.005.

[15] A. Esuli dan F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," hlm. 7.

[16] A. Hamouda dan M. Rohaim, "Reviews Classification Using SentiWordNet Lexicon," hlm. 5.

[17] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, dan C.-J. Lin, "LIBLINEAR: A Library for Large Linear Classification," hlm. 31.

[18] F. Nurifan, R. Sarno, Institut Teknologi Sepuluh Nopember, dan K. Sungkono, "Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMoWikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle," *IJIES*, vol. 12, no. 6, hlm. 47–58, Des 2019, doi: 10.22266/ijies2019.1231.05.

[19] Suhariyanto, A. Firmanto, dan R. Sarno, "Prediction of Movie Sentiment Based on Reviews and Score on Rotten Tomatoes Using SentiWordnet," dalam *2018 International Seminar on Application for Technology of Information and Communication*, Semarang, Sep 2018, hlm. 202–206. doi: 10.1109/ISEMANTIC.2018.8549704.

[20] B. Y. Pratama dan R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," dalam *2015 International Conference on Data and Software*

Engineering (ICoDSE), Yogyakarta, Indonesia, Nov 2015, hlm. 170–174. doi: 10.1109/ICODSE.2015.7436992.

[21] M. Fikri dan R. Sarno, "A comparative study of sentiment analysis using SVM and SentiWordNet," *IJEECS*, vol. 13, no. 3, hlm. 902, Mar 2019, doi: 10.11591/ijeecs.v13.i3.pp902-909.

[22] Z. Wang dan X. Xue, *Support Vector Machines Applications*. Springer, 2014. [Daring]. Tersedia pada: https://link.springer.com/chapter/10.1007%2F978-3-319-02300-7_2

[23] H. T.K, "Random decision forest," *International Conference on Document Analysis and Recognition ICDAR*, hlm. 278–282, 1995.

[24] L. H. Suadaa dan A. Purwarianti, "Combination of Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF) in Indonesian text clustering with labeling," dalam *2016 4th International Conference on Information and Communication Technology (ICoICT)*, Bandung, Indonesia, Mei 2016, hlm. 1–6. doi: 10.1109/ICoICT.2016.7571885.