Sentiment Analysis and Topic Modelling for Identification of Government Service Satisfaction

Moh. Nasrul Aziz, Ari Firmanto, A. Miftah Fajrin, R. V. Hari Ginardi
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
nasrul.17051@mhs.its.ac.id, ari.17051@mhs.its.ac.id, miftah.17051@mhs.its.ac.id, hari@its.ac.id

***Abstract*— The era of information disclosure and social media tends to make people express their opinions on social media. Indonesia is one of the top five nations social media users in general, especially Twitter. This becomes an interesting thing when trying to see public opinion on a government service. Opinion mining can be used to get information from textual twitter to be processed into an information by classifying existing information into positive information classes and negative information classes. In this research, we try to do opinion mining on public opinion about Identification card (KTP) service in Surabaya city. We compare between supervised and unsupervised methods to see their performance for each classifier. In unsupervised the sentiwordnet approach is used to classify between negative and positive opinions. Supervised Support Vector Machine (SVM) method is used to create a classification model to define an opinion. Before the data is classified, preprocessing steps are used to make the data better. In addition, the Latent Dirichlet Allocation (LDA) approach is used to see for topics that tend to be strong which affects a negative or positive opinion. The result of the classification model by using SVM achieved accuracy rate of 75%.**

***keywords*—text mining, sentiment analysis, opinion mining, big data analysis, lexicon based, machine learning, topic modeling, lda**

I. INTRODUCTION

In the era of information disclosure, many people tend to communicate their opinions in social media. Various types of social media development and contribute to the formation of big data about various public opinion. May the opinion about a government service make people tend to write their opinions on social media such as Twitter. From data published by one of Paris-based research institutes, semiocast, Indonesia is the country with the fifth largest user account twitter in the world and ranked third in the activity of posting to twitter [1]. According to [2] Opinion mining is the process of automatically understanding, extracting, and processing textual data for information. Based on the definition of opinion mining can be used to obtain information from textual twitter to be processed into an information by classifying existing information into positive information classes and negative information classes. Machine Learning Based Approaches [3] [4] uses several machine learning algorithms (supervised or unmanned algorithms) to classify data. Lexicon Based Approaches [5] [6] uses dictionaries containing both positive and negative words to determine the polarity of sentiment. Hybrid Approaches [7] [8] [9] uses a combination of Machine Learning approaches and lexicon for classification.

In this research, big data analysis on twitter data about public opinion on population service in Surabaya city. The data to be used to analyze public opinion comes from twitter data. After data is collected then use Lexicon Based Approaches method to obtain sentiment polarity from every twitter sentence. The sentiwordnet approach was proposed for this study. Previously some preprocessing methods used to clean the data in advance include: case folding, cleansing, word formalization, and stemming. To compare the method we proposed, we also made a comparison using the Machine Learning Based Approach. We used the SVM method to classify experiments using machine learning. In addition to classification research also focuses on getting a topic that becomes trending in every class of positive and negative opinion. Using LDA this study will determine each topic that becomes a core of each class of opinion. To measure the research performance used confusion matrix to determine true positive, true negative, false positive, and false negative from the proposed classification. Writing is divided into 5 sections: introduction, related work, method, result, and conclusion.

II. RELATED WORK

Pang and Lee [10] conducted preliminary research on machine learning in the classification of text. By using various methods such as Naïve Bayes, Maximum Entropy (ME), and SVM to analyze the film review data obtained from IMDb.com. They experimented with various engineering features, where SVM produced the highest accuracy of 82.9% with the unigrams feature. They concluded that discourse analysis, focus detection, and joint reference resolution can improve accuracy.

Dang et al. [11] sentiment analysis using SVM by applying different feature selection methods. Experiments were carried out on text data containing 305 positive reviews and 307 negative reviews. SVM is trained in three sets of features that are defined based on domain-free features, domain dependencies, and sentiments. Besides that, feature selection is also done using the Information Gain method. The results obtained by applying feature selection have a better accuracy level of 84.15%.

In another experiment Sahu and Ahuja [12] analyzed the movie review dataset on the IMDB dataset by using the target class from the film rating between 0 (very disliked) to 4 (very like). The steps taken in his research are Pre-processing which includes Stemming, Stopping, Part Of

978-1-5386-5529-0/18/$31.00 ©2018 IEEE

Speech Tagging. Then the next step is Feature extraction by using information gain approach and impact analysis feature which aims to find out the impact of each feature on document polarity. Then a ranking is obtained from information gain on all features. Then the Random Forest classification method, Decision Tree, Naive Bayes, K-Nearest Neighbor were used to measure the success rate of the proposed feature extraction. The result is the extraction feature that they propose has a good impact on the classification success. The best results are obtained using the Random Forest method with an accuracy rate of 88.95%.

Shinjee et al. [13] proposed an integrated topic model based on grouping of TV users and recommendations for the same TV program for social TV services. The proposed model combines two latent Dirichlet allocation (LDA) bound together with the parameters of the proportion of topics: the topic of TV users (groups) and the topic descriptions of words (groups) of the TV programs being watched. To make TV program recommendations, identification of semantic relationships between TV user groups and TV program descriptions was conducted. From the TV user topic model, users with the same tastes can be classified as topics for the recommendations of the social TV community. The study used data obtained from the history of TV shows and TV program guides by TV polling agencies. The experimental results showed that the proposed model produced 81.4% precision for 50 topics on TV program recommendations and had a 6.5% higher performance than the topic models from TV users only. For predictions of TV users with new TV programs, the accuracy of predictions is 79.6%.

III. Method

To do this research, we divided several stages of research. The first stage is the collection of data taken from twitter through the API which we then store in a database management system (DBMS). The next step to improve the classification of course the need for preprocessing data that has a step. The steps are case folding, cleansing, word formalization, and stemming. Subsequently, a classification process was conducted with Lexicon Based Approaches and Machine Learning Based Approaches.

In the Lexicon Based Approaches approach is used sentiwordnet to determine the polarity of each twit user. While on Machine Learning Based Approach used Term Frequency and Invers Document Frequency (TF-IDF) for feature extraction which is then used for training and testing data by using 10 folds cross validation. After the polarity is determined based on sentiwordnet we try clustering by using LDA to search for each of the most popular topics for each class of opinion. For more details on the research, Figure 1 will explain the flow of this research.

A. Data Collection

Data collection step started with the data crawling from Application User Interface (API) which has been provided by twitter through the page https://api.twitter.com. The twitter API serves as a link between systems that was built with twitter. The twitter API requires consumer key, consumer access, access token and access secret tokens obtained by registering the twitter API app at http://dev.twitter.com. After the request via the twitter API will return data with JSON format, then the data stored in a DBMS for processing in the next section.

B. Data Preprocessing

The purpose of the initial processing is to prepare the text into data that will further processing. Text mining requires some initial section which in essence is preparing for the text to be changed to be more structured. In this research preprocessing stages are :

- **Case folding** is the stage that changes all the words in the document into lowercase.

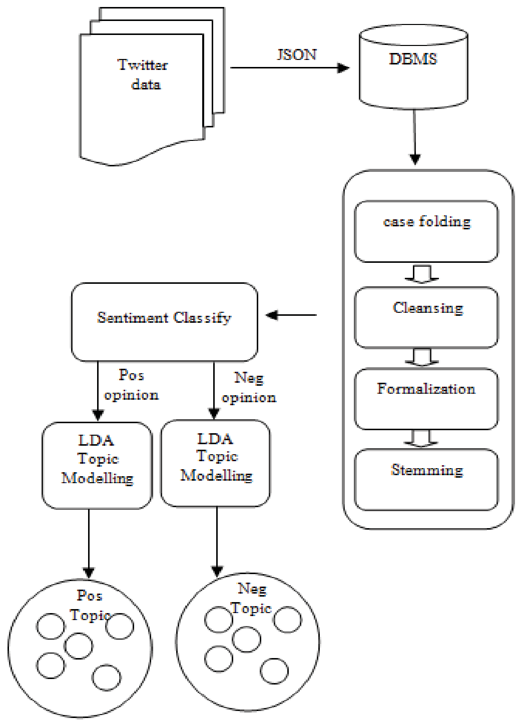

Fig. 1. Proposed Framework

- **Cleansing** required to extract the characters used in this study. In this stage we remove punctuation, symbols, emoticons, user mentions, and hashtags. The characters allowed in this stage are only letters.
- **Formalization:** Twitter data is limited to 160 characters only. This causes the tendency of twitter users to write non-standard sentences. To solve non-standard word problems a word formalization application is required to embed the word in its default form. In this study we used a library from InaNLP [14] which converted the words that are not standard into standard word according to EYD Bahasa Indonesia.
- **Stemming** that is process which contained in a word processing system that transforms words contained in a document to its basic words by using certain rules. In stemming the Indonesian language is different from the English language. In the English language generally to remove word suffix. But in stemming the Indonesian language is not enough to eliminate the suffix only, but also prefix and suffix. This study uses the algorithm Nazief and Adriani [15] to perform the process of stemming in the Indonesian language.

C. Sentiment Classify

The polarity of the document is calculated from the polarity of the sentence. Therefore, the document was first dissected into a sentence. Then, for every sentence, the polarity is determined by using SentiWordNet. Each word is derived and marked, stops the word deleted. If the word originates from one of the class words "adjective", "verb" or "noun", it is searched in SentiWordNet. Scores of corresponding synset are collected. To determine the polarity of each twitter data is to classify it based on Lexicon. SentiWordNet approach is done to recognize the polarity of each data. Each word is derived and marked, stops the word deleted. If the word originates from one of the class words "adjective", "verb" or "noun", it is searched in SentiWordNet. Then the score of every word occurrence is calculated. Then the sentiment score can be calculated using equation (1) to calculate the positive score, the negative score can be calculated with Equation (2) and Equation (3) is used to calculate objectivity [16].

$$Score_{pos} = \frac{1}{n} \sum_{i=1}^{n} Score_{pos}(i) \quad (1)$$

$$Score_{neg} = \frac{1}{n} \sum_{i=1}^{n} Score_{neg}(i) \quad (2)$$

$$Score_{obj} = \frac{1}{n} \sum_{i=1}^{n} Score_{obj}(i) \quad (3)$$

The mentioned procedure has three scores for each twitter data is calculated the polarity. To determine whether the sentence entered in its class is used a comparison to see which is maximum among the three scores (heading, neg, obj) with Eq. (4) to find the maximum value.

$$Polarity = MAX(Score_{pos}, Score_{neg}, Score_{obj}) \quad (4)$$

Using the maximum value equation we can determine the largest value of each score and then determine the polarity.

D. Rule Based Extraction

In the writing of sentences on twitter is not uncommon people who write sentences with the use of words that tend to be positive but has a real meaning negative. Rule based checks by creating rules to see if the sentence contains negation, or question word. In a sentence containing negation, we assume that the original negative sentence may turn positive and negative, as in the sentence "*pengurusan ektp tidak lama*". In the sentence is not meaningful negative, long also meaningful negative when in fact "*tidak lama*" is positive. Some examples of negation in Indonesian are: "*belum, bukan, tak, tanpa, tidak, pantang, jangan, bukanlah, sok, tidak pernah*".

In addition, the rule making to see whether the sentence contains a question or not. In a sentence that contains a question word like "*apakah, bagimana, berapa*" will have meanings that tend to be neutral, because it is just a question.

E. Topic Modelling

Latent Dirichlet Allocation is one of the techniques in topic modeling. The basic intuition of the LDA is a generative probabilistic model of a set of text data (corpus). The basic idea of LDA is that documents can be represented as mixed models of various topics *p* (*z*) and each topic is a series of mixed *p* (*w* | *z*) mixed words. Thus a word in the text can be deduced as Eq. (5) [17].

$$P(w_i) = \sum_{j=1}^{t} P(w_i \mid z_i = j)\, P(z_i = j) \quad (5)$$

To generate the word with LDA can be seen in Figure 2 [18]. The first stage of the alpha-ultra-parameter distribution of Dirichlet extracts the relationship between the document and the theta-d topic, theta-d is a T-dimensional vector, T represents the number of topics in the document. Each element in theta-d represents the possibility of a topic appearing in the document. Second, extract the z-dn topic for the current word from Multi-distribution, and the parameter for Multi-distribution is theta-d. Finally, the original word w-dn was obtained from Multi-phi-z distribution. Through the above procedure we can obtain a combined probability distribution for all words by topic, the concluded procedure can be seen in Equation (6).

$$P(w, z|\alpha, \beta) = P(w|z, \beta)\, P(z|\alpha) = \int P(z|\theta)\, P(\theta|\alpha) d\theta \int P(w|z, \varphi)\, P(\varphi|\beta) d\varphi \quad (6)$$

From Equation (6), we can see the combined probability distribution for all words with two-part topic. One is the probability distribution of words below the topic and the other is the probability distribution of topics under the document. Through this two-part computation we get a combined probability distribution for all words with the topic.

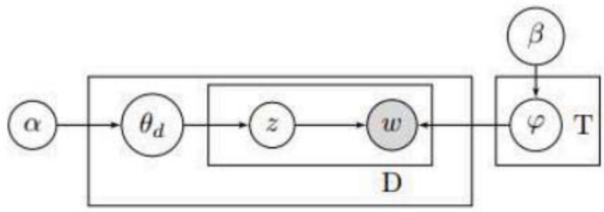

Fig. 2. LDA Modelling

To choose the number of topics, this research used topic coherence model [18]. For coherence measurement, C_V is used that combines the indirect cosine measure with the NPMI (Normalized Pointwise Mutual Information) and the boolean sliding window. The equation can be seen in equation 7.

$$C_v = (S_{set}^{one}, P_{sw(110)}, m_{cos(nlr,1)}, \sigma_a) \quad (7)$$

Where S_{set}^{one} is the segmentation of word subsets that compare one word with the whole words. $P_{sw(110)}$ is the probability of sliding window that using 110 window size. $m_{cos(nlr,1)}$ is indirect cosine measure using NPMI as vector space of words. And $\sigma_a$is the aggregate of coherence score using arithmetic mean.

IV. Result

This study uses data obtained from twitter, total data used amounted to 370 tweet data. In our data collection we use some keywords such as *ektp surabaya*, *ktp surabaya*, *ktp* *service*, etc. The collected data does not have the target class of opinion. To make the target class opinion used simple voting method of three people. Everyone analyzes whether the sentence has a positive, neutral or negative opinion. From the initial data collected data obtained the distribution of classes that can be seen in Figure 3.

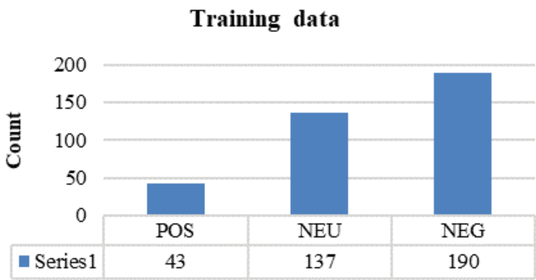

Fig. 3. Training Data

From Figure 3, the initial distribution of data to be used as test data and training data on SVM obtained 43 positive opinions, 137 neutral opinions, and 190 positive opinions. Furthermore, to measure the performance of the classification used the test by forming a confusion matrix with 10 folds cross validation. Then we will be calculated True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). of the results are used to find accuracy (7), Recall (8), and F1-score (9).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F1-Score = \frac{FP}{FP+TN} \quad (9)$$

A. Unsupervised

This step used sentiwordnet method for the classification from the dataset. The opinions obtained from the lexicon based and the results of the method can be seen in Figure 4.

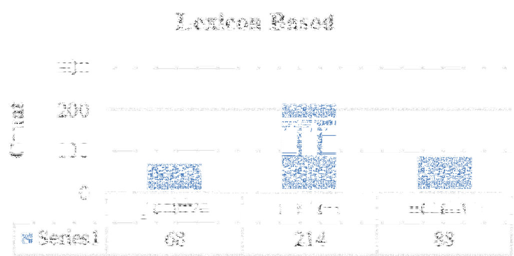

Fig. 4. Result of Lexicon Based

Figure 4 shows that the neutral opinion equal to 214, negative opinion equal to 88 opinion, and positive equal to 68 opinion. From result of applying unsupervised method with sentiwordnet, the test applied to see performance of classification model by accuracy, recall, and f1-score. The results of classification testing can be seen in Table 1.

TABLE 1. PERFORMANCE UNSUPERVISED

	Result		
	Accuracy (%)	Recall (%)	F1-Score (%)
Negative	68	32	44
Netral	40	59	48
Positive	28	51	37
AVG	**53**	**45**	**45**

The experiment obtained accuracy, recall and F1-Score with values of 53%, 45%, and 45% respectively. The use of various languages, especially java language on twitter affect the level of accuracy obtained by the classifier. The system we developed is still not able to process the word in the Java language that is widely used by residents of Surabaya. Therefore, in the next experiment the use of Java language needs to be considered in the data processing to increase the level of classification accuracy.

B. Supervised

Our second experiment used supervised machine learning. SVM with polynomial kernel was chosen to build a classification model to separate the opinions of each twitter sentence. After doing the preprocessing data, the next step is to change all datasets into the word vector. TF-IDF is used to form the attributes used in the classification process. Data is divided into 10 parts to develop test scenarios with 10 times cross validation. Table 2 shows the confusion matrix of classification performance test results with machine learning

TABLE 2. SVM CONFUSSION MATRIX

	Negative	Neutral	Positive
Negative	178	11	1
Netral	67	65	5
Positive	9	6	28

Based on Table 2, we can calculate the accuracy, recall and F1-Score as in Table 3 on supervised performance. In the second experiment with supervised machine learning approach, a higher level of classification performance was compared with the use of lexicon based. This can be seen from the accuracy of 75% generated by SVM classification. In our case, using machine learning has better performance than lexicon based.

TABLE 3. PERFORMANCE SUPERVISED

	Result		
	Accuracy (%)	Recall (%)	F1-Score (%)
Negative	70	94	80
Netral	79	47	59
Positive	83	65	73
AVG	**75**	**73**	**72**

C. Topic Modelling

After the process of determining the sentiment, then the next step is to find the topic in each positive and negative sentiments. In this research, topic coherence model used to find appropriate topic for each negative and positive sentiments. The coherence score can be seen in figure 5 and figure 6.

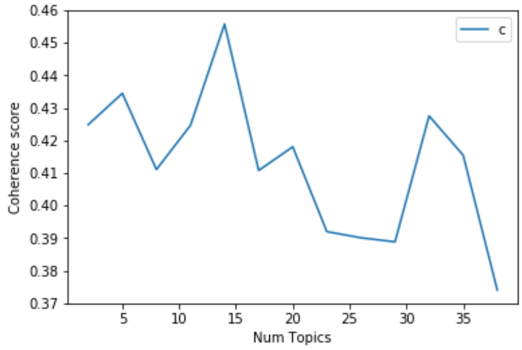

Fig. 5. Topic Coherence Score of Negative Sentiment

From figure 5 and 6, the number topic with highest coherence score is 14 for negative sentiments and 11 for positive sentiments. To analysis LDA results, Termite is used in this research. Termite visualize topic model based on saliency term and term probabilities [19]. The circle area represented term probabilities, the wider area the higher probabilities. Termite results can be seen in figure 7 and 8.

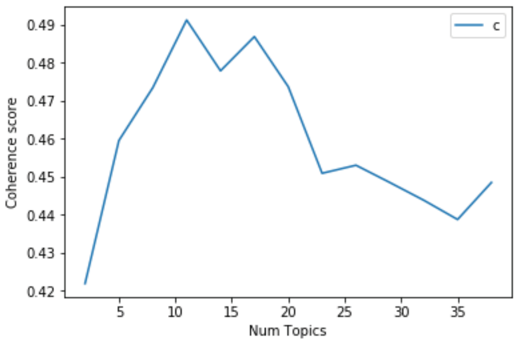

Fig. 6. Topic Coherence Score of Positive Sentiment

From figure 7 and 8, top 5 topics chosen based on the higher of summation term probabilities. The chosen topic has been highlighted in the figure. For the negative sentiment can be analyzed as follows:

1. Topic 0 residents complain about tools that used to record user id card has been damaged because there are keywords *rekam*, *alat*, and *habis*.
2. Topic 7 residents dissapointed that their id card is just sheet of paper after waiting so long to get it. The keyword that represent this topic *hasil, kertas, cetak,* and *data*. For instance of tweet using this keyword is "*emang e ktp ganti kertas hvs tolong tindak terimakasi*h".
3. Topic 8 residents complain about process of creation user id card that takes so long. There are keywords *menunggu, bikin,* and dispenduk, this keyword comes from the tweet "*Setelah menunggu 5 tahun baru dapat ektp*".
4. Topic 10 residents complain about bureaucracy process that very complicated. Complicated means that some residents must go to several government offices to fulfilled the requirements. The keywords are *ektp, urus, susah* and the tweets for instance is "*aju cetak camat ktnya nunggu dispenduk, ektp udah aja sulit camat suruh siola sampe siola suruh camat*".
5. For Topic 13 there are ambiguity words. The keywords with the higher probability terms are *e_ktp, ktp*, and *e* can have a lot of meaning. But they also have small term probabilities using keywords *kertas, siola, and hasil*. So this topic talks about the result of residents id card that just sheet of paper same with topic 7.

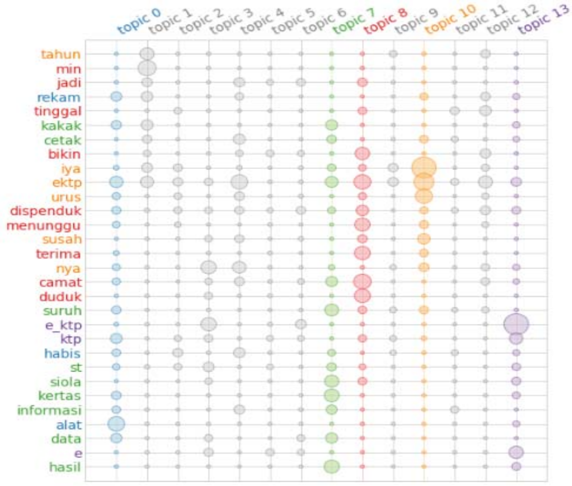

Fig. 7. LDA Result for Negative Sentiment Topics

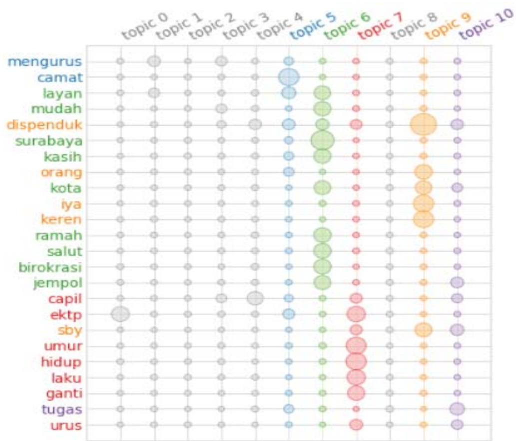

Fig. 8. LDA Result for Positive Sentiment Topics

From the above analyzed about negative sentiments, many residents complain about the awaiting process of id card to finish is so long, the result of temporary id card is just sheet of paper, the complicated of bureaucracy process, and tools used for recording data on user id card is damaged. As for the topic of positive sentiment, can be analyzed as follows:

1. Topic 5 residents pleased with the service from government states. The keywords to express this are *camat, mengurus, layan*. For instance tweet that contain this keyword is "*yang tyt camat dispenduk buka layan sabtuhingga jam siang*".
2. Topic 6 residents pleased with the convenience of government bureaucracy. The keywords to express this are *layan, mudah, surabaya, kasih, kota, ramah, salut, birokrasi, jempol*.
3. Topic 7 residents pleased about the lifetime status of their id card and no need for renewal. The keywords to express this are *capil, ektp, umur hidup, laku, ganti,* and *urus*.

4. Topic 9 residents pleased with the employee of the government. The keywords to express this are *dispenduk, orang, iya, keren*.
5. Topic 10 same with topic 6 and 5, pleased with the services and duty of the government officer. The keywords to express this are *tugas, sby, kota, jempol*.

On the topic of positive sentiment, the residents is very pleased that government serve them until Saturday, the smooth process of bureaucracy, and the lifetime status of thei new id card. This result can be concluded that there are improvement efforts from governemnt to improve its service to the residents

V. CONCLUSSION AND FUTURE WORK

This research developed a method to detect service satisfaction at the government in Surabaya city by utilizing big data analysis. The unsupervised and supervised classification approach is used to find the best method. To handle the less clean data from Twitter, before the classification phase is done preprocessing stage. To define a topic that tends to be strongly in favor of opinion, the LDA is used to model the topics that appear. From the data we collected, we took samples of twitter data relating to population services, especially residents id card. The data used in this study amounted to 370 Twitter data.

The results of this study found that the supervised approach gave more satisfactory results with a comparison of accuracy rate of 75% for supervised and 53% for unsupervised methods. Use of feature extraction using TF-IDF and SVM classification is able to build a better model than the lexicon-based approach. After the classification to find opinions of each data, the next step is to bring up a topic that has a significant effect on positive and negative opinions. The results obtained are 5 topics that dominate negative opinions and 5 topics that dominate positive opinion. Topics that appear and dominate every opinion can be used as advice to the city government of Surabaya as an improvement of service to the community.

Subsequent work on this research can be developed by handling the use of Java language that dominates on twitter data of the people of Surabaya which we consider to improve the performance of classification. In addition, the addition of POS-Tagger feature can also narrow the search space in raising more relevant topics. The use of feature selection for the selection of relevant features on the TF-IDF is considered to improve classification performance and also improve the time efficiency of machine learning. Classification testing by the hybrid method using the lexicon-based score as a feature in the classification is also possible to see the level of performance classification generated.

For the topic of modeling, there are several things that can be taken as an evaluation material from Surabaya city office of population as the authorities. The first is about the tools used in the process of making an id card, the awaiting process of id card to finish is so long, the result of temporary id card is just sheet of paper and the complicated of bureaucracy process. However, some residents also gave positive aspirations such as the opening of service is dispatched on Saturday.

REFERENCES

[1] N. W. S. Saraswati, "Text Mining dengan Metode Naïve Bayes Classifier dan Support Vector Machine untuk Sentimen Analysis," Program pasca Sarjana Universitas Udayana Thesis, 2011.

[2] Pang, B., and Lee, L., “Opinion mining and sentiment analysis. Foundations and trends in information retrieval,” 2(1-2), 1-135, 2008

[3] J. Bollen, H. Mao, X. Zeng, “Twitter mood predicts the stock market,” J. Comput. Sci., 2, pp. 1-8, 2012

[4] Y.-M. Li, Y.-L. Shiu, "A diffusion mechanism for social advertising over microblogs," Decis. Support Syst., 54, pp. 9-22, 2012

[5] Z.-H. Deng, K.-H. Luo, H.-L. Yu, "A study of supervised term weighting scheme for sentiment analysis," Expert Syst. Appl., 41, pp. 3506-3513, 2014

[6] H.N.T. Xuan, A.C. Le, L.M. Nguyen, "Linguistic features for subjectivity classification," International Conference on Asian Language Processing (IALP), November 2012, IEEE, pp. 17-20, 2012

[7] A. Abbasi, S. France, Z. Zhang, H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Trans. Knowl. Data Eng., 2 (3), 2011

[8] Aliaksei Severyn, Alessandro Moschitti, Olga Uryupina, Barbara Plank, Katja Filippova, "Multi-lingual opinion mining on YouTube," Information Processing & Management, Volume 52, Issue 1, Pages 46-60, 2016

[9] D. Ekawati and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, Denpasar, pp. 1-6, 2017

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up: Sentiment classification using machine learning techniques," in Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10, ser. EMNLP '02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79-86.

[11] Y. Dang, Y. Zhang, H. Chen, "A lexicon-enhanced method for sentiment classification: an experiment on online product reviews," IEE Intell. Syst., 25 (4) (2010), pp. 46-53

[12] T. P. Sahu and S. Ahuja, "Sentiment analysis of movie reviews: A study on feature selection & classification algorithms," 2016 International Conference on Microelectronics, Computing and Communications (MicroCom), Durgapur, 2016, pp. 1-6

[13] S. Pyo, E. Kim and M. kim, "LDA-Based Unified Topic Modeling for Similar TV User Grouping and TV Program Recommendation," in IEEE Transactions on Cybernetics, vol. 45, no. 8, pp. 1476-1490, Aug. 2015.

[14] A. Purwarianti, A. Andhika, A. F. Wicaksono, I. Afif and F. Ferdian, "InaNLP: Indonesia natural language processing toolkit, case study: Complaint tweet classification," *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, George Town, 2016, pp. 1-5

[15] Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S.M., & Williams, H.E. Stemming Indonesian: A confix-stripping approach. ACM Trans. Asian Lang. Inf. Process., 2007, pp. 1-33.

[16] K. Denecke, "Using SentiWordNet for multilingual sentiment analysis," 2008 IEEE 24th International Conference on Data Engineering Workshop, Cancun, 2008, pp. 507-512.

[17] T. Pan, W. Zhang, Z. Wang and L. Xu, "Recommendations Based on LDA Topic Model in Android Applications," 2016 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), Vienna, 2016, pp. 151-158.

[18] Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM ’15* (pp. 399–408).

[19] Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization Techniques for Assessing Textual Topic Models. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI ’12*, 74.