

Sentiment Classification of Microblogging in Indonesia Airline Services using Support Vector Machine

Tien Rahayu Tulili
Informatics Engineering
Politeknik Negeri Samarinda
Samarinda, Indonesia
tien.tulili@polnes.ac.id

Muhammad Farman Andrijasa
Informatics Engineering
Politeknik Negeri Samarinda
Samarinda, Indonesia
andrijasa@gmail.com

Abstract— Twitter, as one of microblogs, has become a predominant media used in expanding the service quality to the customer. In Indonesia, for example, most of National airline companies employ Twitter as a media in order to notify their customers about up-to-date information as well as to monitor all real time comments from the customers. Varied opinions (positive, negative, or netral) expressed by netizens through micro-blogging are interesting to analyze. Analysis of netizens' responses to the services provided by the companies can be done with a study of Sentiment Analysis. In this recent study, data collection technique used was crawling technique available on Crawler4j. Furthermore, some processes would be performed to convert unstructured data into structured data including cleaning, case folding, parsing and filtering. The data were then extracted into a certain group of words and the words were then simplified into their basic form using Vega Algorithm. Each word's weight will be calculated by using the TF-IDF method. Each comment then was classified into positive, negative, or netral opinion with Support Vector Machine method. Two open source library, those were LIBSVM and LIBLINEAR used in the experiment. The results obtained that the text can be classified with the accuracy of 93.7128%.

Keywords—Support Vector Machine; sentiment classification; opinion; micro-blogging;

I. INTRODUCTION

Various responses - both positive and negative responses - expressed by the public can be a parameter to measure the extent to which the companies have provided services needed by the public, so that the companies' managements can evaluate their services and plan appropriate strategies to improve their services and attractiveness persuading the public to use them.

Based on survey results issued by PT. Generasi Digital Internasional (GDILab) on November 3, 2016 [1], the top five airline companies widely discussed and exposed on Twitter are Garuda Indonesia, Citilink, Sriwijaya, AirAsia, and Lion Air. From the results of this survey, it is not known how many comments or positive, negative, or neutral responses about the flight services. The various responses expressed by netizens on Twitter can be analyzed. The analysis of netizens' responses to company services can be done using Sentiment Analysis study.

In this analysis, important data from Twitter relating to corporate services both in the field and in social media will be involved and then normalized and classified by certain methods so as to produce useful output.

From the above-mentioned background, some of issues that will be resolved in this study are how netizens' responses via micro-blogging - especially Twitter - to airline services are classified into positive, negative, or neutral responses using Support Vector Machine (SVM) method.

This study aims at applying Support Vector Machine method in the classification process of public opinion to several national airlines and analyzing the sentiments on Twitter. Furthermore, in this research the experiment in classification used two different libraries, those were LIBSVM and LIBLINEAR implementing four kernels and three solver types respectively, and these libraries would be compared to see which of these would give better performance.

II. ABBREVIATIONS

Some abbreviations are used in this research, such as LIBSVM and LIBLINEAR. These two term refers to two different functions available in Weka application. LIBSVM is an open software used for Support Vector Machine[13]. LIBLINEAR is an open source library used for classfying large instances and features [14] [15]. There are some types of kernel and solver types used in the research. RBF stands for radial basis function implemented in LIBSVM. Furthermore, several solver type used in LIBLINEAR experiment, namely L2R_L1LOSS_SVC_DUAL standing for L2-Regularized L1-Loss Support Vector classification; L1R_L2LOSS_SVC standing for L1-Regularized L2-Loss Support Vector Machine; and L2R_LR is L2-Regularized Logistic Regression.

III. RELATED WORKS

One of studies have been conducted related to the analysis of sentiments was a research conducted by [1]. In this study, the classification methods used were the Naive Bayes Classifier

(NBC) and the Support Vector Machine (SVM). The data analyzed were in the form of Indonesian language opinions taken from Indonesian rubric and English opinions taken from an English film review. The results of this research showed that SVM method has better performance than NBC in classifying texts of opinion both in Indonesian and English. Similarly, a research conducted by [2] on English online mass media, the results obtained showed that the accuracy level given by the SVM method was better than that of the NBC method. The result of this research is SVM method has better performance than NBC method to classify text of opinion both in Indonesian and English. Similarly, research conducted by [2] on online mass media. Furthermore, [3] conducted analysis of sentiments toward public figures in Indonesia using NBC method. In their research, an application was built to visualize the popularity of public figures. Still based on the same data source topic that is public figure, [4] in their research did analyzing sentiments toward political elites using the Additive Kernel SVM method namely Intersection Kernel and Polynomial Degree 2. The results of the test showed that Intersection Kernel gave an accuracy of 90% with a comparison of training data and test data of 90:10. In a study conducted by [5], sentiment analysis was done by taking the data source in the form of Indonesian news articles from websites. Methods used in the research were SVM and K-Nearest Neighbors (KNN). The results of the research showed that the test done using k-fold cross validation with the value of k was 10 showed the most accurate validation result of the two methods. In addition, although the processing time required by SVM method was longer than that by KNN method, SVM method still showed a better accuracy than KNN method.

Several studies have used Twitter as a source of data related to sentiment analysis with various classification methods [6, 3, 7, 8, 9]. There were also researches using a combination of two social media Twitter and Facebook [10].

According to Liu [11], sentiment analysis or also called with opinion mining is a field of study that analyzes opinions, sentiments, evaluations, judgments, attitudes, public emotions on entities such as products, services, organizations, individuals, issues, events and other attributes. Opinions, sentiments, judgments, and etc. are expressed in text form. The text may be a review, blog, discussion, news, comment, feedback, or other similar documents. Furthermore, the term of opinion mining appeared first in the writings of [12] According to them, an opinion mining tools will process the search results of a particular item, generate a list of attributes of a product (quality, features, etc.), and sum the opinions grouped in good, bad, or both categories.

IV. APPROACH

The data was collected from Twitter page with the retrieval time ranging from March 16th to May 18th, 2017 from several national airline services Twitter accounts, among others: Lion

Air, Garuda Indonesia, Batik Air, Sriwijaya Air, Citilink and Wings Air. Twitter accounts used as the retrieval parameter value include: @lionairid, @lionairgroup, @indonesiaGaruda, @batikair, @batikairina, @sriwijayaair, and @citilink. Data were retrieved using the functions in Crawler4j. Every tweet taken from the Twitter page, the data taken consists of the time of retrieval in the forms of date and time when the tweet is published, the Twitter identity is also called the Twitter ID, and the content of the tweet. All data are grouped by airline types in separate files.

In the next phase, all tags such as '<', '>', '/' and so on. in each files were removed and all letters were changed into lowercase letters. In this research, a database was used to reserve the comments which were group into its domain type namely twitterID, total words of comments, content of the comments, and class of the comments. A table of the database used to ease comment filtering and labeling. In this research only, a record and a comment has same meaning and is used interchangeably.

The labeling divided into three classes namely positive, netral, and negatif class was done on word and document level. The criterias of how each comment was labeled to each class are as follow. First, all words which have positive meaning such as "memuaskan", "bagus", "cepat", etc. would be categorized into class positive. Further, a comment contained a positive opinion, for example, "Pelayanannya selalu tepat waktu" was also labeled to class positive. Second, if a comment contain a negative meaning word, it was categorized to class negative. Some examples of negative words are "jelek", "lambat", "tunda", etc. Similar to document level of analysis, a whole comment meaning a negative opinion would be labeled with class negative. Some examples of negative meaning comment were "Gimana sih management @Citilink dan @CGK_AP2 informasi delay or reschedule penerbangan kok ga di infokan Ga professional bgt bikinkeki"; "Kondisi antrian chek in @lionairgroup di Terminal 1 B Maskapai HEBAT @AngkasaPura_2 @kemenhub151 apa tindakan". Last, a comment which has netral words or a netral meaning were labeled with class netral. Some examples of class netral were "Ada yang bisa kami bantu", "mohon bantuannya untuk laptop saya yang ketinggalan". Each comment which could not be grouped into one of each classes was removed. All 9909 tweets taken from the table were filtered and 6954 records were ready to be used in the stemming phase. In the stemming phase, each word is stemmed with Vega Algorithm. Before the classification phase, each word was weighted with TF-IDF method. The number of features resulted in the weighting was 1679 features. All filterings and labelings were done by human.

TABLE 1. Measurement of Precision, Recall, and F-Measure of

CLASS	RBF			Polynomial			Li	
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall
Neutral	77%	100%	87.6%	77.4%	100%	87.3%	83.2%	95%
Positive	0%	0%	0%	0%	0%	0%	37%	17%
Negative	0%	0%	0%	0%	0%	0%	52%	31%

V. SUPPORT VECTOR MACHINE

In this research, there were two different function of SVM used, those were LIBSVM and LIBLINEAR. In LIBSVM, four types of kernel were used to search which of those type would give the best result. Those types are RBF, Polynomial, Linear, and Sigmoid. Furthermore, in LIBLINEAR functions, there were three different solver types used, those were L2R_L1LOSS_SVC_DUAL, L1R_L2LOSS_SVC, and L2R_LR_DUAL. All functions were available in Weka application[16].

As explained before that after labeling, extration, TF-IDF calculation, and matrix creation, each pair of matrix consisting TF-IDF value, features, and class was classified based on the classifier using RBF, Polynomial, Sigmoid, and Linier kernel. The formulation of each kernel can be seen further in [13]. Furthermore, the formulation of each solver type can be seen further in [14]. Figure 1 depicts the process of how data was classified.

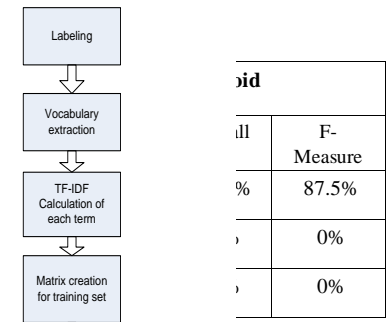


TABLE 3 Results of

LIB- LINEAR	LIBSVM library (kernels)		LIBLINEAR Library (Solver Type)	
	RBF	Polynomial	Sigmoid	Linier
Accuracy	93.7128%	93.2196%	93.0758%	

Fig. 1. Support Vector Machine Classification

VI. RESULTS AND EVALUATION

Before classifying, the records were separated to two groups of data—training data and testing data. There were two scenarios used in this experiment, those are scenario 1, which used ratio 70:30 – 70 percent of the total documents for training and 30 percent of the total documents for testing, and scenario 2, which used ratio 60:40 – 60 percent for training and 40 percent for testing. For classification, SVM method was used to classify each record to class neutral, positive or negative. For the ease of the research, Weka[16] was used and several available functions implemented SVM method, such as LIBSVM and LIBLINEAR, were implemented to classify each record. Since LIBSVM has different kernel type, each type of kernel was implemented in the experiments as well as LIBLINEAR with each of its solver types. Each scenario which used all functions mentioned in the previous and each results was shown on the tables.

For scenario 1, the result can be seen on table 1. First experiment used function LIBSVM. As can be seen from table 1 that the most accuracy result of data classification was

TABLE 4. Results of experiments of scenario 2 using LIBSVM

LIBSVM	Kernel Type (C=1, Gamma=0)			
	RBF	Polynomial	Linear	Sigmoid
Accuracy	79.4393	79.2955	78.6485	79.2955

TABLE 5. Measurement of Precision, Recall, and F-Measure of LIBLINEAR Scenario 1

LIB- LINEAR	L2R_L1LOSS_SVC_DUAL			L1R_L2LOSS_SVC			L2R_LR_DUAL		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Neutral	93.2%	98.7%	96.1%	93%	98.7%	95.8%	92.8%	98.8%	95.7%
Positive	95.2%	57.7%	71.9%	91.9%	58.1%	71.2%	93.9%	56.3%	70.3%
Negative	93.6%	82.7%	87.8%	94.7%	79.8%	86.6%	94.8%	78.9%	86.1%

LIBSVM using Linear kernel type, that was 78.1629%. The other three types of the kernel has given the same accuracy, that was about 77% each.

Furthermore, to give more details of how the performance of each type of kernels can be seen on table 2. Overall, Linier has given the best performance in precision in each category of document classification, 83.2 %, 37%, and 52% for class neutral, positive, and negative respectively. While 100% of documents can be classified as neutral document with RBF, Polynomial, and Sigmoid kernel, only 91% of documents successfully classified as neutral document with Linear kernel. However, using Linear at least 38.3% and 17.9% of documents can be classified as either negative or positive document, compared to other types of kernel which has given none of classified documents. Furthermore, for each F-Measure measure of each type of kernel, RBF is the best performance in classifying neutral documents. But, for classifying positive and negative documents, Linier type has given best performance.

96.1%, 71.9%, and 87.8% of documents correctly classified as neutral, positive, and negative class.

For scenario 2, the dataset was separated into 60% of the total documents for training dataset and 40% of the total for testing dataset. It can be seen from table 4 that in scenario 2, the kernel type which gave the best accuracy was RBF kernel, that is 79.4393%.

Furthermore, the details of each kernel type in this scenario can be seen on table 6. Starting with the F-Measure, the performance of LIBSVM using kernel type RBF, Polynomial, and Sigmoid particularly in classifying documents to class neutral have given better performance, that was 88.5% each, than that of using Linier which was 85.5%. However, in classifying documents to both class positive and negative, Linear kernel type has given the best performance, those were 17.9% and 38.3% respectively, compare to the rest kernel type, was 0% each. Moreover, the performance of LIBSVM using Linear kernel shows that this type of kernel has given better

TABLE 6. Measurement of Precision, Recall, and F-Measure of LIBSVM of scenario 2

LIBSV M	RBF			Polynomial			Linear			Sigmoid		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Neutral	79%	100%	88.5%	79.3%	100%	88.5%	85.5%	89.2%	87.2%	79.3%	100%	88.5%
Positive	0%	0%	0%	0%	0%	0%	29.1%	16.4%	21%	0%	0%	0%
Negative	100%	1%	2%	0%	0%	0%	49.4%	46%	47.6%	0%	0%	0%

As can be seen from table 3, classification using LIBLINEAR function has given better performance. The table shows that L2R_L1LOSS_SVC_DUAL give the best accuracy to classify all documents, which is 93.7128% of documents successfully classified correctly.

Moreover, the performance of each solver type can be seen on table 5. Overall, the result shows that all of the types have give satisfaction performance, that is above 90%. Beginning with the F-Measure, L2R_L1LOSS_SVC_DUAL has the best performance in classification in each of class documents.

performance in classifying documents to both class positive and negative, those were 29.1% and 49.1% (precesion

TABLE 7 Result of experiments of scenario 2 using LIBLINEAR

LIB- LINEAR	Solver type (C=1, bias=1)		
	L2R_L1LOSS_SVC_DUAL	L1R_L2LOSS_SVC	L2R_LR_DUAL
Accuracy	77.6779%	78.6125%	79.7628%

measure) for each class, than the other kernel types with the exception RBF kernel type in classifying negative documents, that was 100%. Regarding to recall measures, it was still Linear type which gave the best performance particular in classifying positive and negative documents.

Moving to the table 7, it shows that the best performance of LIBLINEAR function in classifying the documents was that of

used in the experiments particularly in classifying both negative and positive documents did not give best results. Furthermore, labeling the class to each tweet is done on word and document level by human. Labeling process particularly labeling neutral or positive comment need to be considered further since words used in either neutral or positive tweets were quite similar so that when labeling a tweet it needs to judge in a whole sentence not by looking the one-by-one word only. The

TABLE 8. Measurement of Precision, Recall, and F-Measure of LIBLINEAR Scenario 2

LIB- LINEAR	L2R_L1LOSS_SVC_DUAL			L1R_L2LOSS_SVC			L2R_LR_DUAL		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Neutral	85.7	83.3	87	85.3	89.3	87.2	85.1	91.2	88.8
Positive	20.9	15.1	17.6	32.5	16.4	21.8	39.2	13.2	19.7
Negative	47.7	44.8	46.2	48.7	45.5	47.1	51.1	44.1	47.3

using L2R_LR_DUAL, 79.7628% accurate.

Furthermore, the details of each performance of each solver type can be seen from table 8. Beginning with the result of F-Measure, it shows that even though the result of L2R_LR_DUAL's F-measure in classifying positive documents was only 19.7%, the other two results were the highest among other solver types, those were 88.8% and 47.3% for neutral and negative classification. Moreover, the highest result of Precision in classifying neutral documents was that of using L2R_L1LOSS_SVC_DUAL, but the highest results of precision for positive and negative class were those of using L2R_LR_DUAL. In term of Recall, L2R_LR_DUAL has given best performance in classifying neutral documents, but L1R_L2LOSS_SVC showed the best performance in class negative and positive.

Based on the result of scenario 1 and scenario 2 that document classification using SVM has given various accuracies ranging from 77% to 93%. In scenario 1 which separates dataset to 70% for training data and 30% for testing data, the best performance of text classification is function LIBLINEAR which use L2R_L1LOSS_SVC_DUAL solver type. The best result is 93.7128%. Furthermore, in scenario 2 which separates dataset to 60% for training data and 40% for testing data, the result of accuracy is different. The result of the best performance of the same function that is LIBLINEAR is only 79.7628%.

The ratio of number of documents among neutral, positive, and negative documents in dataset are not balanced. The neutral documents dominate all the documents which accounts for 5442 documents, while positive and negative documents are only 380 and 1132 documents each. This unbalanced document numbers caused the performance of the functions

number of features resulted before weighting was about 1679 different words.

VII. CONCLUSION

The best result of the performance of SVM in classifying sentiment on microblogging is 93.7128% with the ratio number is 70:30 for training and test each. Despite this accuracy, there are some important points should be considered such as the usage of grid search to find out the best value of C and gamma; the ratio of numbers of documents used in the classification; the size of each tweet involved in the classification.

For further research, feature extraction would use n-gram and the ratio of numbers of each class documents would be justified to obtain a balanced ratio. In addition, the size of each tweet would be normalized.

REFERENCES

- [1] <https://blog.gdilib.com/infografis-industri-penerbangan-di-twitter-2/?platform=hootsuite>. Accessed on March, 24 2017
- [2] Saraswati, N. W. S. (2013). Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis. Seminar Nasional Sistem Informasi Indonesia, 587–591.
- [3] Hidayat, Andi Nurul. (2015). Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. Jurnal Elektronik Sistim Informasi Dan Komputer (Jesik), 1(1), 1–7.
- [4] Ramdhani, M.A., Rohim, O. N. (2014). Analisis Sentimen Untuk Mengukur Popularitas Tokoh Publik Berdasar Data Pada Media Sosial Twitter Menggunakan Algoritma Datamining Dengan Teknik Klasifikasi. Jurnal Informasi Volume, VI(2), 1–15.
- [5] Merdekawati, N. N. D, Cholissodin, I., P, Hayunning, J., (2014). Analisis Sentimen Pencitraan Elite Politik Berdasarkan Opini Melalui Media Sosial Twitter Menggunakan Metode Additive Kernel SVM.

- [6] Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM). Seminar Nasional Teknologi Dan Komunikasi 2015, 2015(Sentika), 1–8.
- [7] Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 1320–1326.
- [8] Hu, X., Tang, J., Gao, H., & Liu, H. (2013). Unsupervised Sentiment Analysis with Emotional Signals. International Conference on World Wide Web, 607–617.
- [9] Torkildson, M. K., Starbird, K., & Aragon, C. R. (2014). Analysis and visualization of sentiment and emotion on crisis tweets. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8683(2013), 64–67.
- [10] Khan, A. Z., Atique, M., & Thakare, V. (2015). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. International Journal of Electronics, Communication and Soft Computing Science \& Engineering (IJECSCE), (January), 89.
- [11] Nomleni, P., Hariadi, M., Purnama, I. K. E. (2014). Sentiment Analysis Berbasis Big Data. Rekayasa Teknologi Industri Dan Informasi, 142–149.
- [12] Liu, B. (2012) Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, May.
- [13] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. Accessed on September 2017
- [14] Fan, Rong-En., Chang, Kai-Wei., Hsieh, Cho-Jui., Wang, Xiang-Rui., Lin, Chih-Jen. (2008) LIBLINEAR: A Library for Large Linear Classification. Journal of Machine Learning Research 9, 1871-1874.
- [15] <https://github.com/bwaldvogel/liblinear-java>. Accessed on September 2017
- [16] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.