Classification Methods on Sentiment Analysis of Tourists on Airlines n Twitter

Elza Fitriana Saraswita [a,1], Dian Palupi Rini [b,2,*], Abdiansah [b,3]

[a]Student a Master of Informatics Engineering, Universitas Sriwijaya
[b]Lecturer, Department Informatics Engineering, Faculty of Computer Science, Sriwijaya University,
Jl. Srijaya Negara Bukit Besar, Palembang 30128, Indonesia
[1] 09042621923002@student.unsri.ac.id; [2] dprini@unsri.ac.id*; [3] abdiansah@unsri.ac.id
* corresponding author

ARTICLE INFO

ABSTRACT

Sentiment analysis is one of the knowledge to find the opinions of society towards a topic of discussion particular. Text mining is the science that many performed by individuals or companies to improve performance and fix complaints public against the services or brand trademarks that exist in the world of business. One of them dalah business flight or airline flights. One of them is public complaints against certain airlines posted on twitter. It is certainly going to greatly affect the airline 's own because , media social is one of the means of advertising and trade are extensive. Machine learning methods such as Logistics Regression, Kneighbors Classifier, Support Vector Classifier (SVC), Decision Tree Classifier, Random Forest Classifier, and Gaussian. Several classification methods are used to compare the performance of each method to see the best results .

1. Introduction

All activities are associated with the processing , management and delivery or transfer of information between medium / media when it can be done in the media social[1]. The use of tel e - pound mobile is one of the most much done by the community , especially the use of media social such as twitter, facebook , youtube , google and others . Social media as a place , container or technology facility that removes human boundaries in socializing , space and time boundaries . Utilization of media social in communication , exchange of information , media marketing , trade , look for connections , expand friendship , expressing opinions , exchanging the views and experiences on something things very guarantee . One of the media of social communities popular in giving opinions when it is twitter. Opinions that may be a sentiment which represents private person in conveying their aspirations .

A sentiment about the travel experience of a tourist who is traveling using an airline is one of the parameters to increase business in the world of airlines[2]. This development is supported by the increasing number of social network users who share their opinions and experiences when using a particular airline . A traveler to share his experience in perjanlanan using media social as Twitter to promote the airline low pitch call . Usually that is shared by the rating encompasses comfort , the facilities , the accuracy of the schedule of flights and service of the airline . It is used as an evaluation of competition in the world of business airline flights .

Sentiment is the opinion or view that is based on feelings , emotions , reaction of a person to a case in KBBI . Sentiment analysis is a system of collection and examination of opinions about the products or services that are made in a variety of media social[3]. The use of sentiment analysis can be applied to the various opinions as opinions hatred , opinions general , brand trademarks , and

other media social . Things such as a sign that sentiment analysis can be used to get the message or information specific of the users networking social on a topic that is discussed . Information This can be classified into the several classes for example is very positive , positive , neutral , negative , and very negative.

A few grouping that often times used by researchers to perform analyzes sentiment in social media. Like using a supervised learning approach to conduct sentiment analysis[4] . Researchers trying to use multiple methods of classification such as Logistic Regression, Kneighbors Classifier, Support Vector Classifier (SVC), Decision Tree Classifier, Random Forest classifier and Gaussian for observing , predicting , and comparing the movement of business between the airline and the airline that others.

2. Theoretical Basis

A. Sentiment Analysis

Sentiment analysis is the process of using text analytics to obtain various data sources from the internet and various social media platforms . Interest sentiment antaralain to obtain the opinions of users who are on a platform such. Sentiment analysis acts as a tool that can show a person's emotions that are reflected in writing[5]. Sentiment has 2 categories that differentiate between subjective and objective sentences. The analysis sentiment level is divided into 3 parts, namely message level, sentence level and aspect level. Opinions are divided into categories of explicit opinion and implicit opinion[6]. Part of this is the view or opinions private person it is intended for the case that implied, giving feedback, suggestions, criticism or expressions of hatred. Opinions were expressed in an honest , unequivocal and straightforward and clear or opinions that's are not clear. With that, the message that discussed the topic of a particular make personal or company can obtain input core of users or consumers as efficiently.

B. Sentiment Approach

1. Approach unsupervised or supervised learning are some algorithms that regular use and rely on the data of training[7]. Model classification based on the data trainer who has been given the label in a single domain, often perform poorly with a domain that is different. Although adaptation domain has been studied by the researchers, but the technology is still far from perfect.

2. Approach Lexicon-based is semantic approach which is an approach to using the dictionary sentiment that contain words of opinion and compare it with the data to identify the value of a word[8]. In the Lexicon dictionary , words are compared with their polarity words .

C. Sentiment Classification Method

Classification is one of the main topics in data mining or machine learning [9]. Classification is a grouping of data where the data is used that has a class label or a target. So the algorithms to resolve the problem of classification categorized into the supervised learning or learning that is being watched. Supervised learning approaches that are commonly used for supervised learning approaches , including KNN, Decision tree, logistic regression and support vector. Supervised learning relies on training data. The classification model is based on training data that has been labeled in one domain .

3. Methodology

A. Collecting Data

Dataset is taken from Crowd flower's data is shared in common to anyone just in the form of CSV . Sentiment analysis is about the problem every major airline in United Statet[10]. Twitter data retrieved from February 2015 . Tweet collected as many as 14.848 Data in *Fig. 1* and *Fig. 2* with parameter that exist in the data as many as 15 columns are tweet_id, airline_sentiment, airline_sentiment_confidence, negativereason, negativereason_confidence, airline, airline_sentiment_gold , name, negativereason_gold , retweet_count , text, tweet_coord , tweet_created, tweet_location, and user_timezone . Consists of 3 types of sentiments positive , neutral, and negative.

Fig. 1. Image of US airlines data set
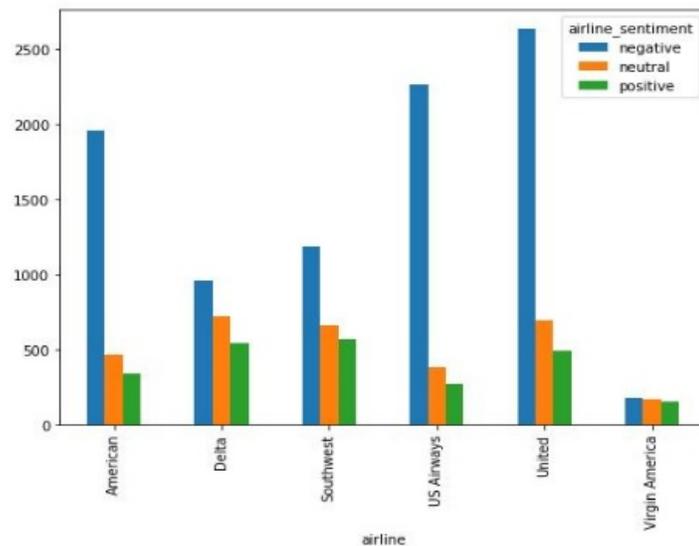
Fig. 2. Figure Graph of US Airlines sentiment tweet count

B. Data Processing

Sentiment analysis has several stages before it can be categorized or classified. Stages sentiment in general can be seen in *Fig. 3*. Text preprocessing stage are early in preparing the text which is

unstructured into data well and ready to be processed[11]. Based on the irregularity structure of the data text, then the system retrieval back information or text mining requires several stages beginning which in essence is to prepare so that the text can be changed into a more structured .
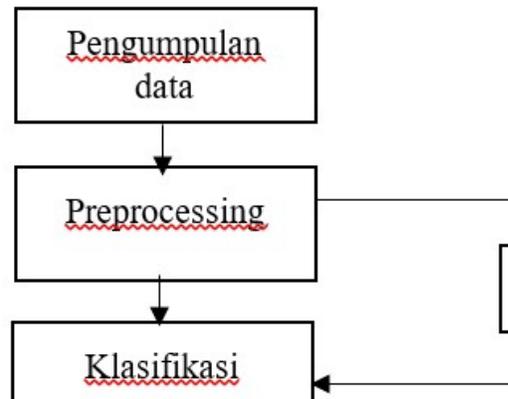
Fig. 3. Figure stages of sentiment analysis in general

1. Pre-processing Stage
   Pre-processing is a technique of data mining that involves changes to the data of crude into a data structure that can be processed . Data crude that the new course taken is often not complete, not consistent, and it contains many errors . There are several stages in pre-processing techniques such as tokenization, case folding, filtering and stemming. Pro-processing techniques are proven to solve these problems .

2. Feature Extraction Stage
   Feature Extraction can be carried out after the dataset becomes more good and structured . This stage is used to get the features that exist in a text so that it will be used as a sentiment classification . On stage is the first to do is perform retrieval characteristics of an object that can describe the objects are . Several types of extraction features include unigram , bigram , and N-gram .

3. Classification Stage
   Classification or grouping is a technique in classifying the data based on the data samples. Researchers try to use multiple methods of classification on sentiment measure the performance of the grouping of several methods that. Among other methods were used by the researchers is Logistic Regression, K-neighbors Classifier, Support Vector Classifier (SVC), Decision Tree Classifier, Random Forest classifier and Gaussian.

4. Evaluation Stage
   In this section testing can be done to get the results of the classification that has been done . Evaluation of these describe what that exist in the data set before experiencing the stages of the processing of the data can be seen in **Fig. *4*** and **Fig. *5***.
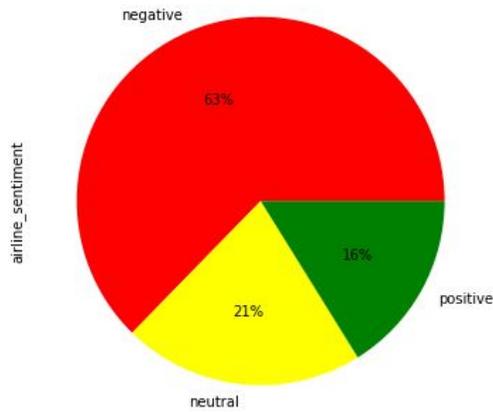
Fig. 4. Pictures percentage sentiment generated by the data set
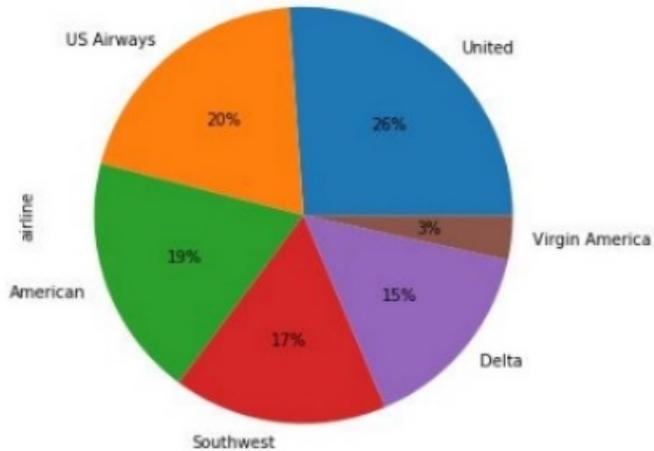
Fig. 5. Pictures of the airline presentation that is much tweeted

5. RESULT

The result of the best of several methods calcification that researchers are trying to do is described in the figure 6 below. The results obtained are Logistic Regression 0.63, K- Neighbors Classifier 0.51, SVC 0.63, Decision Tree Classifier 0.69, Random Forest Classifier 0.75, Ada Boost Classifier 0.72, Gaussian NB 0.4, MLP Classifier 0.74. The best results obtained by researchers using the Random forest method of 0.75 percent .
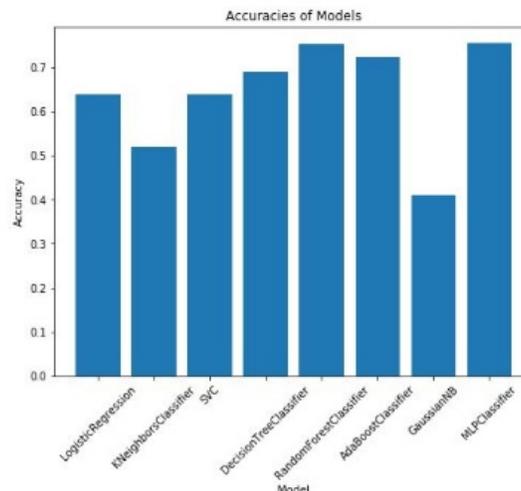
Fig. 6. The classification results

6. Conclusion

Some studies discuss and assess about a method that allows to be able to perform sentiment analysis. Twitter is one of the media of social opinionated people who can support the sentiment analysis to collect datasets. Method approach to supervised or unsupervised use classification (grouping) of the most frequently used even though the results were used not too well or maximum. However , the results were much better can be achieved by improvements in the stage of pre-processing or increase the performance of the method of classification that is used so also with the processing and preparation of the dataset that good .

References

[1]   N. D. Putranti and E. Winarko, "Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 8, no. 1, p. 91, 2014, doi: 10.22146/ijccs.3499.

[2]   Ariyanti, "No 主観的健康感を中心とした在宅高齢者における
健康関連指標に関する共分散構造分析Title," vol. 42, no. 1, pp. 1–10, 2016.

[3]   G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 1, no. 1, pp. 32–41, 2017, [Online]. Available: https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Cal on_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf.

[4]   A. R. Naradhipa and A. Purwarianti, "Sentiment classification for Indonesian message in social media," *Proc. 2011 Int. Conf. Electr. Eng. Informatics, ICEEI 2011*, 2011, doi: 10.1109/ICEEI.2011.6021696.

[5]   A. F. Hidayatullah, "Language tweet characteristics of Indonesian citizens," *Proc. 2015 Int. Conf. Sci. Technol. TICST 2015*, pp. 397–401, 2015, doi: 10.1109/TICST.2015.7369393.

[6]   A. Alamsyah, W. Rahmah, and H. Irawan, "Sentiment analysis based on appraisal theory for marketing intelligence in Indonesia's mobile phone market," *J. Theor. Appl. Inf. Technol.*, vol. 82, no. 2, pp. 335–340, 2015.

[7]   D. T. Alamanda, A. Ramdhani, I. Kania, W. Susilawati, and E. S. Hadi, "Sentiment Analysis Using Text Mining of Indonesia Tourism Reviews via Social Media," *Int. J. Humanit. Arts Soc. Sci.*, vol. 5, no. 2, pp. 72–82, 2019, doi: 10.20469/ijhss.5.10004-2.

[8]   G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," *Citee*, no. June, pp. 39–43, 2014.

[9]   P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 1, p. 145, 2019, doi: 10.24843/mite.2019.v18i01.p22.

[10] E. Susilawati, "Public Services Satisfaction Based on Sentiment Analysis.," *2016 Int. Conf. Inf. Technol. Syst. Innov.*, no. ISBN : 978-1-5090-2449-0., 2016.

[11] M. A. Fauzi, R. F. N. Firmansyah, and T. Afirianto, "Improving sentiment analysis of short informal Indonesian product reviews using synonym based feature expansion," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 16, no. 3, pp. 1345–1350, 2018, doi: 10.12928/TELKOMNIKA.v16i3.7751.