Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 72 (2015) 519 – 526
Procedia Computer Science

The Third Information Systems International Conference

Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers

Nur Azizah Vidya[a], Mohamad Ivan Fanany[b], Indra Budi[b]

[a]*Magister of Information Technology, Faculty of Computer Science, Universitas Indonesia*
[b]*Faculty of Computer Science, Universitas Indonesia, Depok, West Java, Indonesia, 16424*

Abstract

We may see competition among mobile providers to acquire new customers through campaign and advertisement war, especially on social media. The problem arises on how to measure the brand reputation of these providers based on people response on their services quality. This paper addresses this issue by measuring brand reputation based on customer satisfaction through customer's sentiment analysis from Twitter data. Sample model is built and extracted from 10.000 raw Twitter messages data from January to March 2015 of top three mobile providers in Indonesia. We compared several features extractions, algorithms, and the classification schemes. After data cleaning and data balancing, the sentiments are classified and compared using three different algorithms: Naïve Bayes, Support Vector Machine, and Decision Tree classifier method. We measure customer satisfaction on five products: 3G, 4G, Short Messaging, Voice and Internet services. This paper also discusses some correlated business insights in a telecommunication services industry. Based on the overall comparison of these five products, the NBR scores for PT XL Axiata Tbk, PT Telkomsel Tbk, and PT Indosat Tbk are 32.3%, 19.0%, and 10.9% respectively.

Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015)

Keywords: Naïve Bayes; support vector machine; decision tree; sentiment analysis; brand reputation; twitter; Indonesia

1. Introduction

Currently, social media has become more popular among young and senior citizens. Social media users often show their expression by writing their complaints to many objects. In 2014, Indonesia had 20 million active users of Twitter [1]. This number highlights the importance of social media channel for sales campaign of many products and services. Nowadays, social media marketing becoming one of the key strategic brand activities in the world. Good responses from people could awake the desire for a product, create brand awareness, encourage a positive attitude toward the product (brand reputation) and affect intentions for buying the products [11]. Driven by this phenomenon, many telco companies trying to acquire more customers with social media marketing. This paper researches sentiment based brand reputation analysis of three well-known mobile providers in Indonesia. They are PT XL Axiata Tbk, PT

* Corresponding author. Tel.: +62-852-4728-4174.
E-mail address: zeezaah@gmail.com

Peer-review under responsibility of organizing committee of Information Systems International Conference (ISICO2015)
doi:10.1016/j.procs.2015.12.159

Telkomsel Tbk, and PT Indosat Tbk.Currently, telco companies widely uses NPS (Net Promotor Score) to measure their customer loyalty and satisfaction [12]. Socialbakers shows a massive database of Telco social media statistics using NPS [2]. The common claim is that a higher NPS leads to an increase in customer acquisition. However, viewed from social media perspective, Socialbakers shows that some Telco companies have high NPS score but fewer followers. We conjecture that NPS is not accurate enough to evaluate social media campaign. We argue that NPS might be insufficient and limited in its capability in measuring user satisfaction through social media due to following reasons [12, 14]:

- The method used by NPS is aimed only at existing subscribers (current customers).
- NPS is based on the ultimate questions, i.e., by asking recommendation questions to the customers.
- NPS survey is conducted by sampling some customers. The bigger sample size, the more expensive.
- The presented information does not provide detail breakdown analysis on each product.

As mentioned above, we proposed another quantitative measurement to measure people satisfaction by using Net Brand Reputations (NBR) through sentiment analysis. The differences between NPS and NBR methods are shown in Table 1.

Table 1. Method's Differences

Net Promotor Score	Net Brand Reputation
Collecting data by asking customer recommendation (The Ultimate Question).	Collecting data from social media comments (social media listening).
The score simply by existing active subscriber or customer of the brand.	The scores obtained from the entire community without segmentation.
Quantitative measurement with the following parameters: - Promotor (9-10) - Passive (7-8) - Detractor (0-6)	Quantitative measurement from positive and negative confidence from sentiment classification.

2. Related Work

Automatic sentiment analysis through machine learning has been widely studied. Cho et al. proposed a method to visualize the temporal and spatial distribution of brand images using Twitter opinion mining. They build sentiment dictionary for Korean words [14]. This paper showed how Twitter data could be used for brand image analysis across time and locations. Also, the temporal changes in the brand associative network showed which keywords are the focuses of people awareness.

Taysir et al. proposed opinion mining methodology in helping new customers to make a decision about buying or not buying a product by summarizing the reviews [15]. They classified the review's sentences of a product according to features by calculating the cosine similarity. The study ranked features and polarity. The feature classification categorized a class of product using its synonyms. The polarity classification classified sentences into two categories, either positive or negative, according to the polarity of the sentence.

Yu Zhang and Pedro Desouza presented a concept in selecting appropriate classifier based on the features and qualities of data sources. They compare the performances of five classifier with three popular data source in social media: Twitter, Amazon Customer Reviews, and Movie Reviews [4]. They also developed a new sentiment analysis algorithm to enhance the predictive power and accuracy.

Elliot Bricker presented automated sentiment analysis that focuses on analyzing the content of the online post, determining whether they are positive, negative or neutral [17]. The net sentiment score computes the ratio of positive and negative mentions on a topic. NSS helping company to track their brands. Shiv Singh also measures social media influence by identifying net sentiment for some brands

[16]. Mediawave, one of social media analytics in Indonesia, using Net Sentiment for the brand as one of the measurement method on the consumer's loyalty [18].

Along a similar line of research, our study classifies sentiment analysis from Twitter. We build our sentiment dictionary for Bahasa Indonesia and test three classifiers based on naïve Bayes, SVM, and decision tree. We propose a new method to measure brand reputation using Net Brand Reputation, which is similar to Net Promotor Score. We focus on measuring 3G, 4G, Short Messaging Service, Voice, and data or internet. These five services are taken not only because they are common services but because they also give the highest revenue contribution to the telco companies. The score shows a promising result in defining the brand popularity based customer satisfaction and, therefore, defines the best mobile provider to use.

3. Method Description and Sentiment Analysis

A typical workflow of our sentiment analysis can be described as:

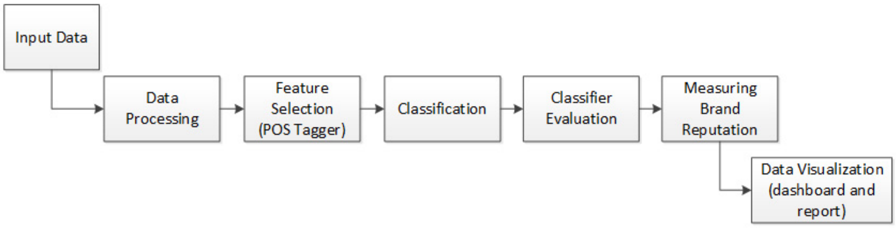


Fig. 1. The workflow of our sentiment analysis

Figure 1 shows overall workflow. Our study uses two basic programming tools. First, we use Rapidminer 5.3 in processing the data up to evaluating classifier after crawling data through Twitter API. Second, we use Tableau 8 to measure brand reputation and automate data visualization into the dashboard. Detailed processes are thoroughly explained in the following seven points.

3.1. Data Description

We obtained the data by writing a script program in PHP that can extract tweets automatically through the Twitter API. Data gathered every end of the week during 12 weeks started from January to March 2015. Data are extracted from the Twitter official campaign account of three mobile providers which are XL Axiata (@XL123), Telkomsel (@telkomsel), and Indosat (@indosatmania).The average number of tweets in every week was around 833 tweets, and giving 10.004 tweets at the end of March 2015. Human annotator labeled each tweet as positive, negative, and unknown. The "unknown" label means human annotator can not understand the tweet. A manual analysis of a random sampling of tweets labeled as "unknown" suggested that many of these tweets contain spams. These spams caused an unbalanced sample of 6237 tweets. We used stratified sampling for each mobile provider to get a balanced dataset of 3000 tweets (1500 tweets each from classes positive and negative).

3.2. Data Processing

Preprocessing is needed to eliminate text noises [4]. It should be performed before text classification process, especially for text that has many non-standard text spelling words [6]. The preprocessing techniques applied in this research are as follows:

- Remove Duplicates

Same tweets or retweet (RT) are removed. This step took data from the last stored tweet that is identified by the oldest number of id_str (unique key).

- Case Folding

All words in the text are converted into lower case.

- URL, RT, punctuation mark, and special character removal

Web address, Retweet (RT), punctuation mark, and special character are removed.

- Removal Stopwords

Words with no particular meaning are removed [8]. Data dictionary using stopwords for Bahasa Indonesia such as 'padahal','akan', etc. and stopwords Twitter such as 'xoxoxox','wkwkw', etc.

3.3. Feature Selection (Part-Of-Speech tagging)

In many previous studies, adjectives, noun, verb, and adverb are important indicators of objectivities and opinions [5]. To define a tag set for Bahasa Indonesia to be used for annotating, a standardized Indonesian annotated corpus is not available. In this study, we used KBBI, which is widely accepted to be the official Indonesian grammar reference [7]. There have been some initial studies conducted on POS tagging for Bahasa Indonesia. Sari et al. applied Brill's transformational rule-based approach in developing a POS tagger for Bahasa Indonesia on a limited tagset. Their tagger was trained on a small manually annotated corpus [5]. This study also used this corpus for POS Tagging tweets with four main parts of speech in Table 2.

Table 2. Data Dictionary POS for Bahasa Indonesia

POS Category	No of Words (Corpus)
kata kerja (Verb)	3.408
kata sifat (Adjective)	19.024
kata benda (Noun)	5.856
kata keterangan (Adverb)	4.162

3.4. Classification

In the modeling and classification step, we use three classifiers algorithm: Naïve Bayes, Support Vector Machine, and Decision Tree. About ten thousands of tweets have been used to build the training set. Manual classification for training set was categorized by five respondents who have a different educational background (informatic, statistic, business management, electro, and communication). The final class of each tweet was determined by the most dominant class chosen by respondents.

3.5. Classifier Evaluation

In the evaluation step, the effectiveness of each classifier was tested with the same data sources. A classifier that gives the best accuracy was used to measure the brand reputation. The accuracy of a sentiment analysis is defined as how well the computer system agrees with the manual classification (human judgment).

3.6. Measuring Brand Reputation

In measuring brand, we used positive mention and then negative mention using Net Brand Reputation index. The NBR is a clean reputation value of a brand digitally, which is understandingly closed to Net Promoter Score (NPS) [6,19,20]. The purpose of this NBR, and also NPS, is to simplify the measurement of consumer's loyalty. The NBR index helps us to focus on creating more positive mentions (promoters) and decreasing the negative remarks (detractors). The NBR can be any number between -100 and 100. The higher the number, the NBR indicates the more positive tweets are considered.

3.7. Data Visualization

To facilitate monitoring and evaluation of brand reputation in real time, we used Tableau 8.2 (analytics tool) by making real-time dashboard. By using Tableau, we can see popular keywords which most frequently discussed, their location, and also the most significant NBR score for each product.

4. Actual Data model and Result

4.1. Algorithms Analysis and Evaluation

The examination of the classification algorithm is based on cross validation technique. We perform some experimental study to found that a fold parameter that could give the best accuracy. This best fold value can be expected to give the best trained model and the lowest error rate [8]. Comparison test result of each fold can be seen in Table 3.

Table 3. Experiment Results to search The Best K for SVM, Naïve Bayes, and Decision Tree

K	2	3	4	5	6	7	8	9	10
SVM	80.70%	80.30%	82.10%	81.50%	82.00%	80.50%	80.70%	80.60%	**82.40%**
Naïve Bayes	77.20%	78.00%	78.80%	78.60%	77.80%	77.51%	77.70%	77.70%	**78.90%**
Decision Tree	72.90%	72.80%	72.70%	72.64%	72.63%	72.56%	72.70%	72.89%	**72.90%**

It could be seen from the result that the best fold parameter for all classification algorithm is 10. The result will be used again in the next step for choosing best classifier algorithm by Precision and Recall score. Comparison of those three algorithms can be seen in Table 4 below.

Table 4. Comparing Performance of Precision, Recall, and F-measure

	Precision	Recall	F-Measure
SVM	86.26%	92.62%	89.33%
Naïve Bayes	91.38%	81.48%	89.08%
Decision Tree	73.05%	100%	84.43%

The experiment on classification algorithm showed that the best algorithm is the Support Vector Machine algorithm with the highest score and lowest different between Precision and recall scores (below 5%). A good score from a model's precision and recall scores can be obtained when both scores are high. Such scores indicate that the model is less possible to be stucked in local minima [3]. We also evaluate the classification algorithms based on the ROC curve (Receiver Operating Characteristic). This curve does not only show the classification accuracy but also visually comparing it with the true positive rate (TPR) and false positive rate (FPR) [8]. The area under the ROC curve (AUC) gives a more comprehensive and gradual classification performance. AUC has a rough guide for classifying the accuracy of a diagnostic test as follows [13]: Accuracy 0.90 - 1.00 = excellent classification, Accuracy 0.80 - 0.90 = good classification, Accuracy 0.70 - 0.80 = fair classification, Accuracy 0.60 - 0.70 = poor classification, Accuracy 0.50 - 0.60 = failure. The result showed SVM had a good classification with AUC score 0.854, whereas naïve Bayes and decision tree had been categorized as failure classification with score 0.5. This score was generated by Rapidminer using performance operator. This result strengthened this study to choose a model formed by SVM algorithm to classify testing data.

4.2. Analysis Data and Measuring Brand Reputation

Classification rules will assign data into positive or negative mentions depending on whether the probability value is higher or lower between two classes. These classes are used as an input parameter in calculating the NBR index. Here we explain our method for calculating the brand reputation of five primary services of each mobile providers. The amount of tweet can be seen in Table 5. Duplicate or spam tweets are labeled by a filter or removed.

Table 5. Total Tweets of Telco Company's Services

	3G	4G	SMS	Voice	Data
	in tweets (1 *tweet* = 140 *characters*)				
XL Axiata	**1,613**	**1,668**	**503**	**1,225**	**2,010**
(-)	544	723	44	461	611
(+)	860	768	454	756	1,143
Filter/remove	209	177	5	8	256
Telkomsel	**1,723**	**1,974**	**522**	**1,277**	**2,090**
(-)	648	977	114	521	801
(+)	996	811	352	714	1,097
Filter/remove	79	186	56	42	192
Indosat	**1,441**	**1,479**	**533**	**1,173**	**2,041**
(-)	462	873	214	434	850
(+)	810	428	319	724	1,143
Filter/remove	169	178	0	15	48

This calculation is applied at each of company's products, which are 3G, 4G, SMS, Voice, and Data Services. The net brand reputation is the net value of the brand reputation measured from social media. This value can also be estimated by the net promoter score [20,21]. In our study, we compute the NBR as the percentage of positive mentions minus the percentage of negative mention as follows:

$$NBR = \left(\frac{Positive\ mentions - Negative\ Mentions}{Positive\ Mentions + Negative\ Mentions}\right)\ x\ 100\% \qquad (1)$$

The purpose of this NBR is the simplification on the measurement of the consumer's loyalty. The index helps us focus on creating more positive mentions (promoters) and decreasing the negative remarks (detractors). We summarize the calculation in graphs shown in Figure 3.

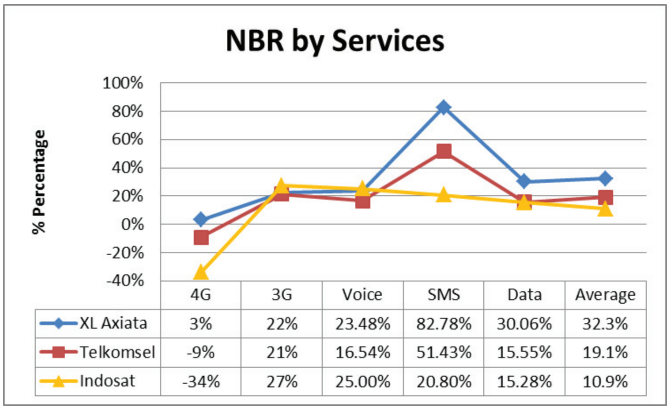
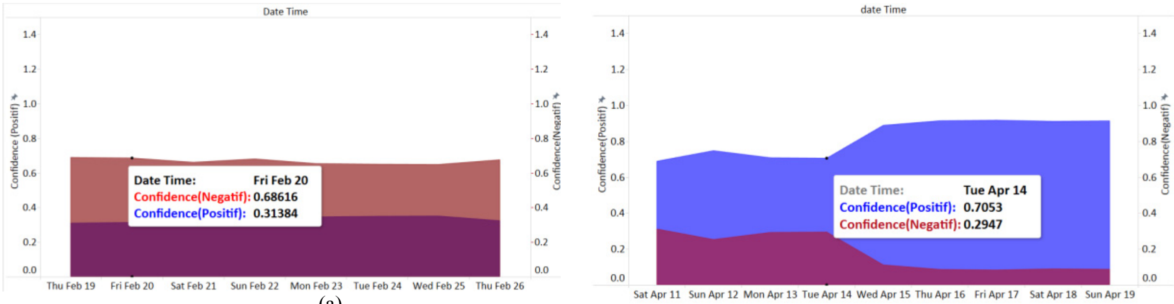


Fig. 2. The overall reputation score of XL Axiata, Telkomsel, and Indosat

Figure 3 shows the percentage of reputation of each service. Positive response dominates the results and do not give a pretty good reputation. The worst reputation comes from 4G. This finding is not surprising for the 4G service is relatively new in all operators (launched at the end 2014). But the worst score comes from Indosat (-34.01%). This fact is might be caused by the larger number of negative responses to Indosat signal and coverage area in providing 4G services to its customer. Indosat was also the last operator who issued 4G services later than other players. The negative response sentiment is also analyzed and shown in the dashboard Figure 4.



(a) (b)

Fig. 3. Probability graph of Sentiment Analysis for (a) 4G services in Indosat and (b) SMS services in XL Axiata

Figure 4(a) shows the average probability distribution for Indosat's 4G services. The graph demonstrates the negative responses dominate comments. The highest NBR scores come from SMS services. It is natural since this services known quite long and stable for all providers. Telkomsel and Indosat give net brand reputation score smaller than XL Axiata, It happened because of the twitter's user comments complained of many advertisements messages received in a day. As well as the dominance of the complaint Telkomsel SMS promo after making calls and sending short messages. XL Axiata has the highest NBR scores. It can be seen from the probability that lead to the positive response during one week in Figure 4(b). This figure shows the probability distribution XL Axiata's SMS services. The graph demonstrates that positive responses dominate comments.

4.3. Analysis Net Promotor Score and Net Brand Reputation

Our Brand reputation calculation by using sentiment analysis found that the Net Brand Reputation scores do not reflect the scores obtained using the Net Promoter Score. There are several methods in data collection and quantitative techniques are different between these two methods, although the purpose is the same, i.e., to analyze public recommendation or public opinion against the company's brand. Comparison results between the NBR and NPS scores can be viewed in Table 6.

Table 6. Score of NBR and NPS

	XL Axiata	Telkomsel	Indosat
NBR	32.3%	19.1%	10.9%
NPS	3	8	3

Table 6 shows that the XL Axiata has a higher reputation score than Telkomsel and Indosat. These results showed that the NPS does not correlate well with the value of NBR, it also highlights the importance of NBR for this method collects data from social media comments. The NBR is quite different from the NPS, which started by providing a direct question to customers on how they recommend the companies brand [9]. On the other hand, NBR technique is also known as social media listening. Social media listening is a social media marketing methods to handle public comments on various social media such as Twitter [10]. Based on an interview, we found that the marketing communication and customer service of XL Axiata seems realized that keeping engagement with social media follower is important. Keeping a mindset which sees them not only as a customer, but also consider them as a friend.

5. Conclusion and Future Work

The results show that the telco companies must have a lot of users who correctly use that services, but they may still do not know public perception on the quality of the services they provide to their customers. Typically, any submitted opinions by users were ignored by the company. From the official Twitter account of XL Axiata (@XL123), Telkomsel (@telkomsel), and Indosat (@indosatmania), we extract opinion and processing that sentiment with three classifier algorithms. The results showed that SVM gives better performance than other two classifiers (Naïve Bayes and Decision Tree) in time processing and accuracy. This results also showed that XL Axiata has a better reputation in five products with average NBR score 32.3%, whereas Telkomsel 19.2% and Indosat 10.9%. Since we made a real-time monitoring dashboard of NBR score, this research can also be applied to various industrial or academic areas through their Twitter official accounts. The dashboard also summarized most keywords that people discussed, thus the company can anticipate or mitigate the possible impact to their brand. For future study, we plan to elaborate our training model by doing domain adaptation or tranfer learning using an unsupervised feature representation learning. We also plan to collect more data from social media campaigns of various industries.

Acknowledgement

This work is supported by Higher Education Center of Excellence Research Grant funded by Indonesia Ministry of Research, Technology and Higher Education. Contract No. 0475/UN2.R12/HKP.05.00/2015.

References

[1] R. Bergevin, A. Kinder, W. Siegel and B. Simpson, Call Centers for Dummies, Mississauga: John Wiley & Sons Canada, Ltd., 2010, p. 345.

[2] Socialbakers, "Twitter Brand Statistic," Socialbakers, 2008. [Online]. Available: http://www.socialbakers.com/statistics/twitter/profiles/brands/telecom/. [Accessed 24 January 2015].

[3] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Factor," pp. 1-22, 2007.

[4] Y. Zhang and P. Desouza, "Enhance the Power of Sentiment Analysis," in *International Journal of Computer, Information, Systems and Control Engineering*, Hopkinton, 2014.

[5] Sari, Syandra and M. Adriani., "Developing Part of Speech Tagger for Bahasa Indonesia Using Brill Tagger," *The Iternational Second Malindo ,* p. 1, 2008.

[6] B. Patel and D. D. Shah, "Significance of stop word elimination in meta search engine," *International Conference On Intelligent Systems and Signal Processing (ISSP,* pp. 52-55, 2013.

[7] H. Alwi, S. Dardjowidjojo, H. Lapoliwa and A. Moeliono, Tata Bahasa Baku Bahasa Indonesia, 3rd ed, Jakarta: Balai Pustaka, 2003.

[8] J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques: Concepts and Techniques third edition, Waltham: Elsevier, 2012.

[9] D. Chaffey, "The Net Promotor Score," in *E-Business & E-Commerce Management fifth edition*, Pearson, 2011, pp. 465-467.

[10] F. Reichheld, Ultimate Question, Harvard Review Business Press, 2006.

[11] D. Evans and j. Mackee, Social Media Marketing: The Next Generation of Business Engagement, Canada: Wiley Publishing, 2010.

[12] T. PT XL Axiata, "Digital Report," XL Internal, Jakarta, 2014.

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

[14] S. W. Cho, M. S. Cha, S. Y. Kim, J. C. Song dan K.-A. Sohn, “Investigating Temporal and Spatial Trends of Brand Images using Twitter Opinion Mining,” *International Conference on Information Science and Applications (ICISA),* 2014.

[15] T. H. A. Soliman, M. A. Elmasry, A. R. Hedar and M. M. Doss, "Utilizing Support Vector Machines in Mining Online," in *International Conference on Computer Theory and Applications*, Alexandria, 2012.

[16] S. Singh, "Applying Metrics to SIM Realm," in *Social Media Marketing For Dummies*, Hoboken, John Wiley & Sons, Inc., 2011, pp. 80-82.

[17] E. Bricker, "Can Social Media Measure Customer Satisfaction?," NetBase Solutions, Inc., Berkeley, 2011.

[18] Mediawave.biz, "PT MediaWave," 14 September 2014. [Online]. Available: http://mediawave.biz/.