

# Using Text Mining to Analyze Mobile Phone Provider Service Quality (Case Study: Social Media Twitter)

Calvin and Johan Setiawan

**Abstract**—Competition between telephone providers to attract new customers can be seen through advertisement war on TVs, posters and radios nearly every moment. Question is arise on how do we measure the quality of these providers in order choose the best one for oneself.

This paper is written to solve the question by measuring customers satisfaction by using text mining. Sample model is extracted from social media Twitter and the sentiment polarity is measured using Naïve Bayes classifier method. The model shows a promising result on defining the popularity based on customer's satisfaction and therefore defining the best provider to be used

**Index Terms**—Naïve bayesian, sentiment analysis, telephone provider, text mining.

## I. INTRODUCTION

Nowadays people use telephone to send messages despite the distances between them. There are many providers and programs available for us to choose from which creates competition between these companies and yet confusion for the users.

Arising along the internet popularity, social media Twitter become one of the top web accessed and used by Indonesian. Millions of tweets containing thoughts, questions, comments and critiques posted daily. The telephone provider companies even use this media to get closer to the customers. These huge amounts of posts [1] can easily become a source of information, of course it has to be polished first.

This paper suggests a method to extract information by using text mining [2] and naive Bayesian method on the model of extracted Twitter posts. Sentiment analysis is also used to identify the readers opinion to determine positiveness of the posts [3]. A few examples also showed how effective Sentiment Analysis such as [4]-[6].

## II. RESEARCH OBJECT AND ASSUMPTION

The telephone providers that are going to be research objects are 3 famous telephone provider in Indonesia. They are PT XL Axiata Tbk, PT Telkomsel Tbk and PT Indosat Tbk

PT XL Axiata is one of the biggest telephone provider company in Indonesia with broad network and high quality service across the country that has stood since October 8<sup>th</sup>

1996. PT XL Axiata was proclaimed to be the first private company that provides telephone services especially for mobile phone in Indonesia.

PT Telkomsel Tbk has been established since 1995 as one of the innovator to develop Indonesia's communication technology. To achieve that, Telkomsel keeps to grow their network rapidly through the country while empowering the community. PT Telkomsel Tbk became the pioneer of mobile telecommunication technologies in Indonesia.

PT Indosat Tbk was established in 1967 as a foreign investment company and started to operate in 1969. In 1980, PT Indosat Tbk became state-owned enterprise which is wholly owned by Indonesia's Government. Until now, the company provides cellular services, international communications and satellite services. Most known services from them are Indosat Mentari and IM3.

To carry on this research, there are a few assumptions and criteria had to be made to make this research doable.

### Assumption 1:

The real data is tweets from Twitter Timeline via LingPipe4Twitter which mentions one of three telephone provider companies in this research within 15 Timeline pages and 100 tweets per pages. 1500 tweets considered to be sufficient as a sample for this research.

### Assumption 2:

Every tweets might contains none or more than one positive and/or negative sentiment word. Tweet with no sentiment word won't affect the result of the research.

### Assumption 3:

Repetition of sentiment word in a tweet won't be counted as the previous word is already calculated.

### Assumption 4:

Calculation for the result is focused on the individual word in every tweet, not per tweet.

### Assumption 5:

The 0 (zero) point score are the quality wanted by user. Negative point shows bad quality and positive point shows good quality.

## III. GENERAL METHOD DESCRIPTION AND SENTIMENT ANALYSIS

Research started by Understanding the Literature, especially related to text mining, Naïve Bayes and Sentiment Analysis. Once having sufficient material, the research continues with Determining Positive and Negative Word along with Data Collection from Twitter using

Manuscript received July 9, 2013; revised December 10, 2013.

Calvin and Johan Setiawan are with the Information System Department, Faculty of Information and Communication Technology, Multimedia Nusantara University, Scientia Boulevard Street, Gading Serpong, Tangerang, Banten-15811, Indonesia (e-mail: calv.axl@gmail.com; johansetiawanumn@gmail.com).

LingPipe4Twitter. LingPipe4Twitter used for this research because it is free to use and also a combination of both open source library. The Determined Words will be inserted to the Words Bank / Dictionary. Result from Data Collection is a .csv file. From the .csv file, research will be followed by Data Processing using the Words Bank / Dictionary, so it can be considered feasible to be used in the next phase of the research.

After processing the data, the results that appear are scored for each company. Scoring results will then be analyzed to achieve the desired results. (See in Fig. 1).

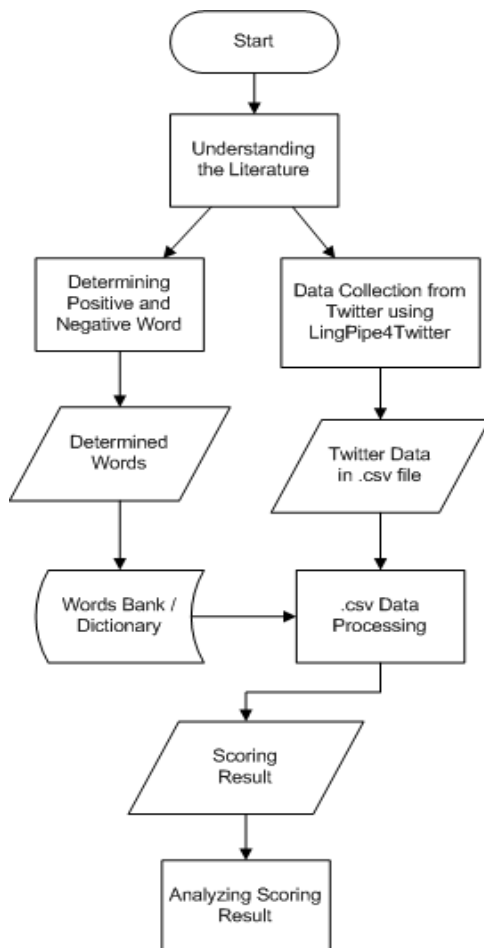


Fig 1. General method flowchart.

Before carrying on to model the data with Naïve Bayes, we need to create word dictionary which will be used as the base of sentiment analysis.

Sentiment Analysis refers to learning an opinion or concerns, feelings and emotions are expressed through writing [7]. The main task in Sentiment Analysis is to classify the polarity of a text in a document, sentence or feature - an opinion in the document, or in a sentence giving positive aspects, negative or neutral. With this, Sentiment Analysis can determine whether the person is in a state of emotion [8].

Sentiment Analysis involves the use of a classification opinion within a text into categories such as "positive" or "negative" and the category is considered "neutral". Sentiment Analysis application will also be able to track what someone is saying about a brand or message [9].

Using Sentiment Analysis, the Determining Positive and Negative Words resulting the Words Bank / Dictionary shown at Table I.

TABLE I: WORDS BANK / DICTIONARY

Words Dictionary	
Positive	Negative
untung	hilang
oke	menyesal
bagus	lemot
terpercaya	lambat
bersahabat	pending
murah	bohong
cinta	jelek
kreatif	mati
lancar	ganggu
hebat	rusak
puas	error
kuat	parah
cepat	menyusahkan
beres	menurun
kencang	berkurang
irit	putus
mantap	tipu
asik	menipu
betah	gagal
aktif	payah
stabil	mengesalkan
jelas	kesal
terimakasih	penipu
jernih	lelet
bebas	berhenti
percaya	masalah
ramah	gangguan
konsisten	mengecewakan
terjangkau	kecewa
senang	menyebalkan
top	susah
hemat	nyangkut
	rugi
	mahal

#### IV. TRAINING SETS

To assure this method is applicable and reliable, it will be tested against a training data model. The training data model is based on the Words Bank / Dictionary to make sure the original data will be processed and shows a correct result.

The training data contains 200 tweets which 100 tweets filled with positive sentiment words and 100 tweets filled with negative sentiment words. For each sentiment word, positive sentiment will be scored as 1 point and negative sentiment will be scored as -1 point. A negation word before a sentiment word will reverse the point.

Here are some of the results of the data training shown in Table II.

Column "articleid" filled with 1(one) shows the positive data training and the field which filled with 2(two) shows the negative data training.

Column "who" shows who made the tweet, whether it is public with positive opinion or the negative opinion. Column "whom" shows whom the tweets mentioned to, in this case is the telephone provider company.

Column "value" shows the point for each word that contains with positive or negatif sentiment/meaning. Column "keyword" and "sentenceprocessed" indicates the words with the following sentence that contains a positive or negative sentiment.

TABLE II: SOME RESULTS ON DATA TRAINING

article id	who	sentence index	whom	value	keyword	sentenceprocessed
1	Public+	14	Tes Positif	-1	untung	ini sih ga untung ganti @XL123 dikit2 pending browsing apapun lambat.
1	Public+	28	Tes Positif	-1	stabil	internet speednya gak stabil ? buffer youtube mpe ngulang 5x @XL123.
1	Public+	38	Tes Positif	-1	jelas	nipunipu nih. @XL123 bnyk potonganpotongan gak jelas pulsa dipotong mulu. Apaapaan nih.
1	Public-	1	Tes Negatif	-1	rugi	koq tiba2 pulsa gw ilang ya Mana nih @XL123 jgn bikin org rugi donk.
1	Public-	2	Tes Negatif	-1	lemot	alah katanya @XL123 oke yg ada malah nyesel udah ga ada sinyal lemot pending pula.
1	Public-	2	Tes Negatif	-1	pending	alah katanya @XL123 oke yg ada malah nyesel udah ga ada sinyal lemot pending pula.
2	Public+	1	Tes Positif	1	untung	wah pake @XL123 untung juga yah ternyata.
2	Public+	2	Tes Positif	1	bagus	oke juga nih @XL123 bagus sinyalnya.
2	Public+	2	Tes Positif	1	oke	oke juga nih @XL123 bagus sinyalnya.
2	Public+	5	Tes Positif	1	bersahabat	terima kasih @XL123 udah ngasih paket2 yang bersahabat ini.
2	Public+	6	Tes Positif	1	murah	oi @tina si @XL123 ada kasih promo murah tuh cobain deh.
2	Public+	7	Tes Positif	1	oke	maju terus @XL123 oke banget!.

## V. ACTUAL DATA MODEL ANALYSIS AND RESULT

The actual data model is obtained from LingPipe4Twitter. Actual data model consists of up to 1500 tweets, where LingPipe4Twitter search the mentioned research's object Twitter account. PT XL Axiata's Twitter accounts are @XL123, @XLCare dan @XLandMe. For PT Telkomsel, the Twitter accounts are @telkomsel, @Kartu\_As dan @simPATI. Then PT Indosat's Twitter accounts are @IndosatMania, @indosat dan @indosatcare.

The search result focusing the number of tweets showed in Table III.

TABLE III: SEARCH RESULT

Company	Account	Search Result	Subtotal
PT XL Axiata	@XL123	1494	4490
	@XLCare	1498	
	@XlandMe	1498	
PT Telkomsel	@telkomsel	1453	4445
	@Kartu_As	1498	
	@simPATI	1494	
PT Indosat	@IndosatMania	1496	4265
	@indosat	1274	
	@indosatcare	1495	
Total Tweets			13200

The search result will be used again in the next step which is Data Processing. Data Processing is a step where scores will be calculated by using a modified Java Application specialized in text mining [10].

The results of this process are shown in the tables below for each company. Table IV is the result for PT XL Axiata, where Table V is the result for PT Telkomsel and Table VI is the result for PT Indosat.

With the total scoring obtained for each company, the result shows that PT XL Axiata Tbk has better service quality

than the other two companies. PT XL Axiata Tbk leads with 29 positive points, followed by PT Telkomsel Tbk with 70 negative values and then PT Indosat Tbk in last position with 100 negative value.

TABLE IV: SCORING RESULT FOR PT XL AXIATA

Account	Tweet	Score	Sub Total
@XL123	Positive	106	34
	Negative	-72	
@XLCare	Positive	65	-29
	Negative	-94	
@XLandMe	Positive	32	24
	Negative	-8	
Total Scoring PT XL Axiata			29

TABLE V: SCORING RESULT FOR PT TELKOMSEL

Account	Tweet	Score	Sub Total
@telkomsel	Positive	95	-59
	Negative	-154	
@Kartu_As	Positive	26	14
	Negative	-12	
@simPATI	Positive	31	-25
	Negative	-56	
Total Scoring PT Telkomsel			-70

TABLE VI: SCORING RESULT FOR PT INDOSAT

Account	Tweet	Score	Sub Total
@IndosatMania	Positive	12	-37
	Negative	-49	
@indosat	Positive	92	-18
	Negative	-110	
@indosatcare	Positive	76	-45
	Negative	-120	
Total Scoring PT Indosat			-100

## VI. CONCLUSION

The result of this research in this paper shows that the

companies in telephone service providers must have had a lot of users, but they may still not know the quality they provide to their customers. Usually any submitted opinions by users were ignored by the company. By processing the opinions that have been submitted using text mining, this paper has shown the service quality of each company.

#### REFERENCES

- [1] S. Iiritano and M. Ruffolo, "Managing the knowledge contained in electronic documents: a clustering method for text mining," in *Proc. the 12<sup>th</sup> International Workshop on Database and Expert Systems Applications*, 2001, pp. 454-458.
- [2] C. Bridge, *Unstructured Data and the 80 Percent Rule*, 2011.
- [3] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and Tomas, "Sentiment analysis on social media," in *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012, pp. 919-926.
- [4] B. Pang and L. Lee. (April 2, 2013). Department of Computer Science, Cornell University. [Online]. Available: <http://www.cs.cornell.edu/home/llee/papers/pang-lee-stars.pdf>
- [5] A. Adrifina, J. U. Putri, and I. W. Simri, "Pemilahan artikel berita dengan text mining," in *Proc. Seminar Ilmiah Nasional Komputer dan Sistem Intelijen, Universitas Gunadarma*, Depok, 2008, pp. 176-181.
- [6] A. Nurani, B. Susanto, and U. Proboyekti, "Implementasi naïve bayes classifier pada program bantu penentuan buku referensi matakuliah," *Jurnal Informatika, Universitas Kristen Duta Wacana*, vol. 3, no. 2, pp. 32-36, 2007.
- [7] B. Liu, *Handbook of Natural Language Processing*, CRC Press, 2010.
- [8] M. D. Haff. (March 12, 2010). Customer think. [Online]. Available: [http://www.customerthink.com/blog/sentiment\\_analysis\\_hard\\_but\\_worth\\_it](http://www.customerthink.com/blog/sentiment_analysis_hard_but_worth_it)
- [9] Lingpipe: Sentiment Analysis Tutorial. [Online]. Available: <http://alias-i.com/lingpipe/demos/tutorial/sentiment/read-me.html>
- [10] Y. E. Soelistio and M. R. S. Surendra, "Simple text mining for sentiment analysis of political figure using naïve bayes classifier method," in *Proc. the 7th International Conference on Information and Communication Technology and Systems*, Bali, 2013.



**Johan Setiawan** was born in Jakarta on October 27, 1964. He graduated from Bina Nusantara University, Jakarta, DKI Jakarta - Indonesia majoring in Information System. He also received his Master of Business Administration degree (MBA) from Monash-Mt Eliza University, Melbourne - Australia, and Magister Management (MM) degree from IPMI - Jakarta - DKI Jakarta majoring in General Management. His primary interests are in data warehouse, data mining and system analysis and design.



**Calvin** was born in Jakarta on January 19, 1992. He graduated from Multimedia Nusantara University, Tangerang, Banten - Indonesia majoring in Information System. His interests are in games and custom application development.