Implementation Text Mining for Recommendation Follow Up Customer

Vivine Nurcahyawati
Information System Department, Faculty of Information Technology
Institut Bisnis dan Informatika Stikom Surabaya
Surabaya, Indonesia
vivine@stikom.edu

Abstract—Customer is one of the biggest assets in a company. The cost of acquiring new customers is greater than the cost of maintaining customer relationships today. The company's follow-up should be appropriate to support customer retention. This study aims to produce applications as a tool to generate recommendations about customer conditions. In this article explained that used a combination of the concept of Mining Text and naïve bayes clasiffier algorithm to process the status of customers from social media, in this study using Facebook. After going through the testing phase, the application can generate recommendation data for follow-up on the customer.

Keywords—Data Mining, Customer Retention, Naïve Bayes Classifier, te

I. INTRODUCTION

In today's increasingly competitive era, organizations have been widely demanded to pay attention to quality (products or services) to their customers. Top leaders and managers are challenged to create and maintain systems and controls to ensure that quality-focused strategies will continue to be implemented and developed [1]. Therefore, an effective organization will give result a good service quality and customer satisfaction.

There are two customer satisfaction strategies: Offensive Strategy and Different Strategy [2]. The offensive strategy is primarily aimed at reaching and acquiring new customers. While difensif strategy includes efforts to reduce the possibility of customer exit and switching customers to other marketers. The purpose of this defensive strategy is to minimize customer turnover and maximize customer retention by protecting its products and markets from competitors' market attacks [2]. If the company is concerned only with offensive strategies and ignoring the strategy difensif, then its survival will be threatened at any time. Because establishing long-term relationships with existing customers will be more effective for corporate growth and increased profitability [3]. Requirements to be met by a company to be successful in the competition is trying to reach the goal of creating and retaining customers [2].

Each company must have its own way in providing services to its customers. Always conduct an evaluation for activities that have been done and always make improvements to the next activity. A wide range of facilities are provided for the convenience of its customers. The belief in the quality of service earned can be a recommendation for new customers or for old customers for subsequent use of services. Providing information on services, facilities, promos and discounts is done on an ongoing basis in an effort to increase the number of subscribers. Customer retention is an important thing for most companies because the cost of acquiring new customers is greater than the cost of maintaining customer relationships today [4]. The contribution of this research is to apply the concept of text mining and combined with Naïve Bayes Classifier (NB) algorithm to classify customers and know the potential of customer retention in order to follow up.

II. IMPLEMENTATION OF TEXT MINING

Text Mining is automatic or semi-automatic processing involving text structures and extracting relevant information on text [5]. Text Mining relates to words that are transcribed and stored in electronic files, representing raw data for analysis. The stages in text mining are as shown in Figure 1.
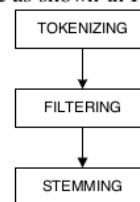

Fig 1. Stage of Text Mining (Miller, 2005)

Text Mining is a process that uses a word approach in conducting its analysis. Text Mining analyzes data in the form of text obtained from sources in the form of documents and the goal is to find words that can represent the contents of the

document so that it can be done linkage analysis between documents. The stages are:

a. Tokenizing stage is the cutting stage of the input string based on each word that compiles it. This process is done to make it easier when done matcing with key words related to the analysis to be done.

b. Filtering is the stage of taking important words from tokenizing process. This process only identifies each incoming word from tokenizing results. The word obtained will be done directly matching process and raised the value of matching results.

c. Stemming Stage is the root searching stage of each word filtering result. In this process the word sorting becomes a basic word without any word affixation.

Customer data used is status data that exist in social media up. Then the data is processed matching with words that come from Bag of Traveling is a set of words related to the term tourism or tourism when doing a trip for recreation or vacation, and also preparations made for activities during traveling (50 miles) from his home with the purpose of recreation, because the term traveling is a definition agreed by the World Tourism Organization [6].

III. USE OF NAÏVE BAYES CLASSIFIER ALGORITHM

Naive Bayes Classifier (NBC) is a statistical classifier that can be used to predict the probability of membership of a class. Naive Bayes is based on the Bayes theorem that has similar classification capabilities to the decision tree and neural network. Naive Bayes proved to have high accuracy and speed when applied into databases with large data [7]. Bayes's prediction is based on Bayes's theorem formulas with the following general formula:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

(1)

The NBC algorithm is applied when calculating the probabilities of stage 2 (filtering) and 3 (stemming) in the text mining stages. NBC performs matching with the status data training so as to produce a probability decision value that the customer data taken included in the category of traveling or spam as in Figure 2.

Fig 2. NBC Flowchart in the Text Mining Process

IV. IMPLEMENTATION OF TEXT MINING AND NAÏVE BAYES CLASSIFIER

There are 2 actors in this system, namely Customer and Marketing Staff. The system retrieves status data created by customers in Facebook's social media. Then the data will be processed by marketing staff. Text mining process is done on the customer's status data by applying the NBC algorithm. Figure 3 describes the functionality of the two actors.
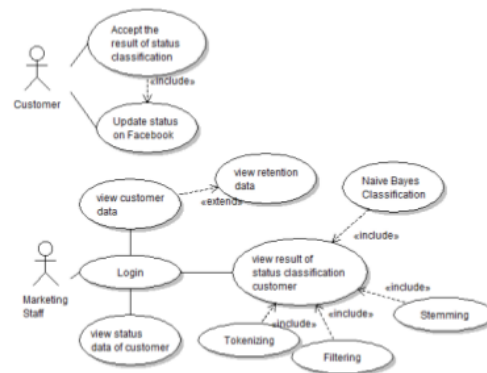
Fig 3. Use Case System Travelling Classification

The collection of customer facebook status is done directly through the application linked to the url of the customer's account. It is seen in Figure 4 that the customer status data will be stored in the database for subsequent processing.

Fig 4. Customer Status from Facebook

The status that has been obtained will be continued to the classification process by using tokenizing, filtering, stemming process in which using NBC algorithm. The result of calculation by using NBC generate weight value whether the customer's status included in the category of traveling or spam. The value of the given treshold is 0.5, meaning that if the weighted value of the NBC calculation exceeds that value it will be categorized as the traveling status. Figure 5 shows weighted results from one customer status. Status is included in the category Traveling.

The classification results are then stored as training data for the system. The more training data the system will run smarter and faster in issuing the results.

Status	Pulang kerja diajakin liburan ke pantai kenjeran. Asyik
Hasil Steamming	pulang kerja diajakin libur ke pantai kenjeran asyik
Hasil Filtering	5
Kata Filtering	libur
Hasil Travel	0.97826086956522
Hasil Spam	0.021739130434783

Fig 5 Point of Status Classification

Figure 6 shows the results of customer classification with the status of traveling. The data is presented by sorting by the highest potential retention value. This data can be used by marketing staff to determine which customers should be given follow-up to the use of company services.

Classification of Travelling Categories Customer

Potential Retention Customer : 6 person

No.	Name	URL Facebook	Potential Point
1	Ainur Kholis Rizkiyanto	https://web.facebook.com/rizkykholis	0.99999988305183
2	Rona Anggraini	https://web.facebook.com/ronaanggraeini	0.99999970399979
3	Caroline Patricia	https://web.facebook.com/carolinepatricia	0.8
4	Teghar Firmansyah	https://web.facebook.com/teghar	0.75
5	Tri Septianto	https://www.facebook.com/tri.septianto	0.75
6	Agil Rijal Quaresma	https://www.facebook.com/agil.quaresma	0.64285714285714

Fig 6. Classification of Travelling Categories Customer

V.   CONCLUSSION

This research implements a text mining process that has tokenizing, filtering and stemming stages. In the process is also combined with Naïve Bayes Classifier algorithm to provide probability value on the classification of customer status data.

The results of the process have been done two kinds of tests, namely using unit testing is also integration and system testing. Unit testing has been performed and the results show that the functionality of tokenizing, filtering, stemming and NBC algorithm usage is in accordance with expected conditions. That means the system to classify customer status is applicable. Integration and system testing tested the status classification process and attempted applications on a number of users.

Based on the research conducted, has produced an application that implements the concept of text mining. Using the app can be used to manage customer status data taken from social media. Management of customer status data as a recommendation material to determine the customer who will follow-up

REFERENCES

[1] Gilbert, G.R. & A.M. Parhizgary. 2000. "Organizational Effectiveness Indicators to Support Service Quality", Managing Service Quality, page: 46-51.
[2] Tjiptono,Fandy , 1997, Strategi Pemasaran, Edisi 1, Penerbit Andi, Yogyakarta.
[3] Fornell, C. (1992). A National Customer Satisfaction Barometer: The Swedish Experience. Journal of Marketing, 56, 6-21.
[4] Singh, Roopa. Khan, Imran Akhtar. 2012. An Approach to Increase Customer Retention and Loyalty in B2C World, International Journal of Scientific and Research Publications ,Volume 2, Issue 6, 1
[5] Miller, T.W. (2005). Data and Text Mining A Business Applications Approach. New Jersey : Pearson Prentice Hall
[6] Holiday, P.G., 2015, Istilah-Istilah dalam Pariwisata. Diambil kembali dari Hotel Book, Trasportation, Tour Operator, Inbound Tour & Travel : http://www.linkedin.com/pulse/istilah-istilah-dalam-pariwisata-pt-great-day-holiday
[7] Eko Prasetyo, Data Mining : Konsep dan Aplikasi menggunakan MATLAB, 1st ed.Yogyakarta,Indonesia: Andi, 2012

Implementation