Análise Descrtiva dos Dados

Aluno: Matheus Raphael Elero

21 de setembro de 2021

1 Introdução

Aqui vai um arquivo com os comandos do "knitr"
Para incluir um pdf, usamos o pacote "pdfpages"

Simulation of Marketplace Customer Satisfaction Analysis Based on Machine Learning Algorithms

Ajeng Aulia Turdjai
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Ganesha Street 10, Bandung 40132, Indonesia
auliajeng@students.itb.ac.id

Kusprasapta Mutijarsa
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Ganesha Street 10, Bandung 40132, Indonesia
kusprasapta.mutijarsa@ itb.ac.id

***Abstract*— Twitter can be a source of public opinion data and sentiment. Such data can be used efficiently for marketing or social studies. In this research addresses this issue by measuring net sentiment based on customer satisfaction through customer's sentiment analysis from Twitter data. Sample model is built and extracted from more than 3.000 raw Twitter messages data from March to April 2016 of top marketplace in Indonesia. We compared several algorithms, and the classification schemes. The sentiments are classified and compared using five different algorithms classification. There are, K-Nearest Neighbor, Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine. The five machine learning can be applied to the Indonesian-language sentiment analysis. Preprocessing on the stages of tokenization, parsing, and stop word deletion of word frequency counting. Frequency of the word of the document used weighted by TF-IDF. The Random Forest, Support Vector Machine, and Logistic Regression generate better accuracy and stable compared with the Naïve Bayes and K-Nearest Neighbor. The results showed Support Vector Machine has accuracy 81.82% with 1000 sampling dataset and 85.4% with 2000 sampling dataset. This shows that the more the number of training data will improve the accuracy of the system. The Net Sentiment score for marketplace in Indonesia is 73%. This results also showed that customer satisfaction has average Net Promoter Score (NPS) 3.3%.**

Keywords*— *Marketplace, Customer Satisfaction, Machine Learning, Algorithms Clasification

I. INTRODUCTION

Based on data from the Socio Economic Status (SES) in 2014, there are 92 million or more than 40% of the bank account linked to a credit card and a debit of Indonesia's population reached 240 million. When compared with the penetration of the mobile phone, the rate is still low because about 85% of Indonesia has a mobile phone where each month they spend 661 pages for browsing [1].

In 2015, We Are Social, a social marketing agency reported on the development of e-commerce based devices used in Indonesia society. There are 18% of active users who accessed the internet through a computer to browse products that may be purchased, and 16% of them purchase goods online at one of the sites accessible. While active users who access the internet via mobile phones have 11% to see the products that may be purchased, while 9% are buying goods online that are accessed through a mobile phone [2].

At this time microblogging site has become a very popular means of communication among internet users. Where millions of messages every day on a popular website that provides microblogging services such as twitter, tumblr, and facebook [3]. This led to more and more users are posting about product and service they use, or express their views on polotics and religion. In 2014, Indonesia has 20 million active users twitter [4]. Twitter can be a source of public opinion data and sentiment. Such data can be used efficiently for marketing or social studies [5].

Twitter sentiment analysis on weaknesses in the words contained in sentences posted by users of the site. Twitter only allows users to write 140 characters, which is why users often use abbreviations and spelling words wrong word. The wrong way of writing resulted in deficiencies in the process of text mining, which can complicate the features taken as well as reduce the accuracy of the classification.

The magnitude of the effect and benefits of Sentiment Analysis, leading research or application of the Sentiment Analysis is growing rapidly, even in America, there are approximately 20-30 companies use Sentiment Analysis to obtain information about public sentiment towards the company's service [6].

Sentiment analysis have influence and benefits, which is to obtain information about public sentiment towards companies [7]. As previous research associated with Sentiment Analysis, among others, research [8] detects fake websites or sites with the original classification of news articles on the website. Research [9] to analyze sentiment on twitter text, using the n-gram language characters and SVM models to cope with high lexical variation in Twitter text. Research [10] developed a system that can identify and classify public sentiment to predict interesting products in marketing.

2 Dados

```
  ind carat     cut color clarity depth table price direcao coord
1   1  0.23   Ideal     E     SI2  61.5    55   326       x  3.95
2   1  0.23   Ideal     E     SI2  61.5    55   326       y  3.98
3   1  0.23   Ideal     E     SI2  61.5    55   326       z  2.43
4   2  0.21 Premium     E     SI1  59.8    61   326       x  3.89
5   2  0.21 Premium     E     SI1  59.8    61   326       y  3.84
6   2  0.21 Premium     E     SI1  59.8    61   326       z  2.31
```

3 Tipo de Variáveis

```
$ind
[1] "integer"

$carat
[1] "numeric"

$cut
[1] "ordered" "factor"

$color
[1] "ordered" "factor"

$clarity
[1] "ordered" "factor"

$depth
[1] "numeric"

$table
[1] "numeric"

$price
[1] "integer"

$x
[1] "numeric"

$y
```

```
[1] "numeric"

$z
[1] "numeric"
```