

INTISARI

Analisis Sentimen menggunakan Metode *Naïve Bayes Classifier* dengan Model Dokumen Bernoulli dan *Support Vector Machine*

Oleh

Tiara Gumilang Ramadhani
11/313042/PA/13626

Berkembang pesatnya sistem informasi berbasis *website* memberikan berbagai kemudahan untuk mendapatkan informasi dalam jumlah besar dan secara cuma – cuma. Informasi yang tersaji dalam bentuk tekstual seringkali berbentuk dokumen yang tidak terstruktur. Untuk menangani dokumen dengan pola tidak terstruktur ini analisis *text mining* sangat diperlukan. Salah satu analisis *text mining* yang sering digunakan adalah klasifikasi sentimen atau klasifikasi teks. Dari hasil klasifikasi sentimen mengenai suatu topik ini dapat disimpulkan opini tertentu dengan melihat proporsi kelas dari keseluruhan dokumen yang disajikan. Metode yang akan dibahas kali ini adalah metode *probabilistic Naïve Bayes Classifier* dan metode *Support Vector Machine* untuk menentukan kelas suatu dokumen secara biner yaitu kelas positif dan kelas negatif.

Naïve Bayes Classifier merupakan metode klasifikasi menggunakan aturan Bayesian dengan memanfaatkan probabilitas prior serta probabilitas bersyarat dari frekuensi kata yang muncul pada masing – masing kelas dokumen *training*. Nilai tersebut yang akan digunakan untuk menentukan kelas dokumen *testing* dengan melihat nilai *maximum a posteriori* masing – masing kelas. Data yang telah diklasifikasi kemudian dihitung tingkat akurasi kebenarannya menggunakan metode *support vector machine* dengan menentukan fungsi Kernel serta proporsi untuk *data training* dan *data testing* yang sesuai. Dari perbandingan nilai akurasi klasifikasi dengan menggunakan kedua metode diatas, didapatkan nilai akurasi paling tinggi adalah dengan menggunakan metode *naïve bayes classifier* (NBC) dengan akurasi sebesar 62,6295%. Dari 5.412 dokumen *training* dan 2701 dokumen *testing* yang telah diklasifikasi, didapatkan proporsi opini untuk kelas positif sebesar 44,46501% dan opini untuk kelas negatif sebesar 55,53499%.

Kata kunci : Klasifikasi sentimen, opini, *text mining*, aturan Bayesian, *Naïve bayes classifier*, *support vector machine*.

ABSTRACT

Sentiment Analysis using Naïve Bayes Classifier with bernoulli Document Model and Support Vector Machine

by

Tiara Gumilang Ramadhani
11/313042/PA/13626

The rise of information system base on website has given a lot of simplicity to get the information in a large number and for free. An information with textual document is usually given as an unstructured form. To dealing these unstructured document form, text mining is being required. One of analysis in text mining which is often used is sentimen classification or text classification. From the result of the classification about one topic we can specify the classes for the opinions by seeing the proportion on each class from the entire documents provided. Two methods that will be discussed are naïve bayes classifier and support vector machine to determine a binary classes for each document.

Naïve bayes classifier is a classification method with Bayesian rule by using the prior probabilities and conditional probabilities in each class from data training. The probabilities is being used to determine the maximum a posteriori each classes for the data testing and specify the document classes. The accuracy for the classified data is being calculated by using support vector machine method with the Kernel function and data proportion determined from the data training. By comparing these two methods, the best accurate rate for the sentimen classification is by using naïve bayes classifier (NBC) which is reached a percentage 62,6295%. By classifying 5.412 data training and 2701 data testing using naïve bayes classifier, the result shows from the entire documents the proportion for the positive class is 44,46501% and for the negative class is 55,53499%.

Keyword : sentimen classification, opinion, text mining, Bayesian rule, Naïve bayes classifier, support vector machine.