

Analisis kinerja pemilihan fitur untuk support vector machine (svm) pada masalah analisis sentimen = Performance analysis of feature optimization for support vector machine svm on sentiment analysis problems / Adhimas Yudha Prawira

Adhimas Yudha Prawira, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20386986&lokasi=lokal>

Abstrak

Twitter merupakan salah satu media sosial yang digunakan secara massive di Indonesia. Para pengguna Twitter ini membicarakan berbagai macam hal, salah satunya terkait pencalonan presiden. Perbincangan para pengguna Twitter ini memiliki nilai sentimen baik positif maupun negatif. Dukungan masyarakat terhadap masing-masing kandidat calon presiden dapat diketahui dengan melihat sentimen masyarakat melalui perbincangan mereka di Twitter, hal ini sering disebut juga sebagai analisis sentimen. Namun, jumlah pengguna dan obrolan para pengguna Twitter yang sangat banyak mengakibatkan data yang akan diproses membutuhkan waktu yang cukup lama. Untuk melakukan proses analisis sentimen para pengguna Twitter secara cepat dan otomatis dapat digunakan bantuan mesin. Salah satu metode yang digunakan untuk melakukan proses analisis sentimen adalah Support Vector Machine (SVM). Pada dasarnya, semakin banyak data yang digunakan sebagai data training dalam pemilihan model fungsi klasifikator maka akan memberikan generalisasi akurasi analisis sentimen untuk data testing yang tinggi pula. Namun di sisi lain, semakin banyaknya data training juga akan menyebabkan besarnya dimensi ruang fitur. Hal ini membuat mesin membutuhkan waktu yang cukup lama dalam melakukan pembentukan fungsi klasifikator. Untuk menanggulangi hal ini, akan dilakukan metode optimasi fitur sehingga mesin dapat tetap membentuk fungsi klasifikator dengan akurasi yang tinggi namun dengan dimensi ruang fitur yang rendah.

<hr>

Twitter is a social media that used in Indonesia massively. Twitter users talk (tweet) about various things, one of them is about presidential nomination. Twitter user conversations have a positive or negative sentiment. Community support for each presidential candidate can be determined by looking at the public sentiment through their conversations on Twitter, this is often referred to sentiment analysis. However, the number of users and tweets cause the data to be processed requires quite a long time. Machine can be used to make the process of Twitter sentiment analysis quickly and automatically. One method that used to perform the sentiment analysis process is a Support Vector Machine (SVM). Basically, the more data that used as data training in the model selection function will give a high accuracy generalization sentiment analysis on data testing. On the other hand, the increasing number of training data will also cause large dimensional feature space. This makes the machine takes a long time to perform model selection. To overcome this problem, feature optimization will be performed. Feature optimization will preserve the high accuracy of the model, but with a low dimensional feature space.