

INTISARI

ANALISIS SENTIMEN COMMIT LOGS DI GITHUB MENGUNAKAN METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES

Oleh

Suci Nur'aini
14/364181/PA/15934

GitHub merupakan salah satu layanan web untuk repositori pengembangan perangkat lunak. Pengembang dapat berbagi *source code*, mengetahui perubahan *source code* dan memberi komentar pada *source code* yang diubah. Pada GitHub terdapat beberapa perintah yang dapat dilakukan oleh pengembang, salah satunya adalah *commit*. Perintah *commit* digunakan untuk mencatat perubahan apa saja yang terjadi pada repositori lokal. Terkadang, pengembang mengekspresikan emosi mereka mengenai kesulitan dalam menangani suatu proyek maupun kepuasan suatu proyek pada *message* tersebut. *Message* tersebut dinamakan *commit logs*. Namun, *commit logs* tidak dapat memberikan kesimpulan sentimen mengenai emosi pengembang secara otomatis, maka diperlukan analisis sentimen. Proses analisis sentimen memerlukan algoritma klasifikasi yang memiliki performa tinggi untuk menghasilkan hasil klasifikasi yang akurat.

Pada penelitian ini metode *Support Vector Machine* dan *Gaussian Naïve Bayes* diimplementasikan untuk membandingkan metode mana yang lebih baik dari kedua metode klasifikasi tersebut untuk kasus analisis sentimen *commit logs* di GitHub. Penelitian ini menggunakan data *commit logs* di GitHub dalam rentang tahun 2007-2013 dengan 70% data sebagai data latih dan 30% data sebagai data uji. Penerapan metode *Support Vector Machine* memiliki performa analisis sentimen *commit logs* di GitHub yang lebih baik dengan akurasi, presisi dan *recall* masing-masing sebesar 0,97 dibandingkan dengan metode *Gaussian Naïve Bayes* dengan akurasi 0,93, *recall* 0,93 dan presisi 0,94. Waktu rata-rata yang dibutuhkan pada proses pelatihan menggunakan metode SVM sebesar 30,872 detik. Sedangkan, pada metode *Gaussian Naïve Bayes* hanya dibutuhkan waktu rata-rata sebesar 0,045 detik.

Kata Kunci: *Analisis Sentimen, GitHub, Support Vector Machine, Gaussian Naïve Bayes*

ABSTRACT

SENTIMENT ANALYSIS OF COMMIT LOGS IN GITHUB WITH SUPPORT VECTOR MACHINE AND NAÏVE BAYES

By

Suci Nur'aini
14/364181/PA/15934

GitHub is one of the web services for software development repositories. Developers can share the source code, know the changes in source code and make comments on the modified source code. On GitHub, there are several commands that can be done by the developer, one of which is commit. The commit command is used to record any changes that occur in the local repository. Sometimes, developers express their emotions about the difficulties in handling a project and the satisfaction of a project on the message. The message is called commit logs. However, commit logs cannot conclude developer sentiments automatically, so sentiment analysis is needed. The sentiment analysis process requires a classification algorithm that has high performance to produce accurate classification results. However, commit logs cannot provide conclusions about sentiments towards developers' emotions automatically, so sentiment analysis is needed. The sentiment analysis process requires a classification algorithm that has high performance to produce accurate classification results.

In this study, the Support Vector Machine and Gaussian Naïve Bayes method are implemented to compare which methods are better between the two classification methods for the case sentiment analysis of logs on GitHub. This research use GitHub's commit logs data from 2007 to 2013 where 70% data as train data and 30% data as test data. SVM method has better performance for GitHub's commit logs sentiment analysis with accuracy, precision, and recall each get 0,97 points whereas Gaussian Naive Bayes method get 0,93 for accuracy, 0,93 for recall, and 0,94 for precision. The average time needed in the training process using SVM method are 30,872 seconds while Gaussian Naive bayes method need 0,045 seconds.

Keyword: *Sentiment Analysis, GitHub, Support Vector Machine, Gaussian Naïve Bayes*