

INTISARI

ALGORITMA NAÏVE BAYES DAN COMPLEMENTARY NAÏVE BAYES PADA HADOOP FRAMEWORK UNTUK ANALISIS SENTIMEN FILM

Oleh :

Rininta Vita Rosanti

12/334747/PA/14978

Saat ini, media sosial menjadi alat komunikasi yang sangat populer diantara pengguna internet. Setiap harinya, jutaan pengguna internet membagikan pemikiran dan opininya terhadap topik tertentu, salah satunya adalah opini mengenai film yang sedang populer. Dari opini dan ulasan film tersebut dapat dilakukan analisis sentimen apakah suatu opini termasuk ke dalam sentimen positif, negatif, atau netral.

Pendekatan analisis sentimen dapat dilakukan dengan menggunakan *machine learning* dan algoritma klasifikasi. Algoritma *Naïve Bayes* dan *Complementary Naïve Bayes* merupakan metode klasifikasi yang menggunakan perhitungan probabilitas.

Pada penelitian ini dilakukan implementasi algoritma *Naïve Bayes* dan algoritma *Complementary Naïve Bayes*, untuk melakukan analisis sentimen dan klasifikasi data *tweet* film berbahasa Inggris. Algoritma ini akan dijalankan di atas *Apache Hadoop Framework* dan menggunakan *Apache Mahout* sebagai *machine learning library*.

Hasil dari penelitian ini menunjukkan bahwa Algoritma *Complementary Naïve Bayes* memiliki performa yang lebih baik dalam melakukan klasifikasi pada data yang tidak seimbang, dengan akurasi 2,57% lebih tinggi, *precision* 1,92% lebih tinggi, *recall* 2,57% lebih tinggi, dan *F-measure* 2,03% lebih tinggi jika dibandingkan dengan algoritma *Naïve Bayes*.

Kata Kunci : analisis sentimen, klasifikasi, *machine learning*, *naïve bayes*, *complementary naïve bayes*, *apache hadoop*, *apache mahout*

ABSTRACT

NAÏVE BAYES AND COMPLEMENTARY NAÏVE BAYES ALGORITHMS ON HADOOP FRAMEWORK FOR MOVIE SENTIMENT ANALYSIS

By :

Rininta Vita Rosanti

12/334747/PA/14978

Social media today has become a very popular communication tool among internet users. Millions of users share opinions on particular topic everyday, for example opinion about the latest popular movie. From those opinions and movie review, sentiment analysis can be done to identify whether the opinion expressed in a text is positive, negative, or neutral.

Sentiment analysis approach is often done by machine learning using the classification algorithm. *Naïve Bayes* and *Complementary Naïve Bayes* algorithms are classification algorithms that uses a probability calculation.

In this research, the implementation of *Naïve Bayes* and *Complementary Naïve Bayes* algorithms are used to perform sentiment analysis and data classification on english movie tweet. Those algorithms will be run on top of *Apache Hadoop Framework* and *Apache Mahout* as machine learning library.

The result from this research conclude that *Complementary Naïve Bayes* algorithm showed better performance than *Naïve Bayes* algorithm within the classification sentiment on skewed data bias, with 2.57% higher accuracy, 1.92% higher precision, 2.57% higher recall, and 2.03% higher F-measure when compared with *Naïve Bayes* algorithm.

Keywords : sentiment analysis, classification, machine learning, *naïve bayes*, *complementary naïve bayes*, *apache hadoop*, *apache mahout*