

Klasifikasi Sentimen Data Ulasan Produk Pakaian Menggunakan Metode *Multinomial Naïve Bayes* Berbasis *Information Gain* dan *Query Expansion Ranking*

Riema Meristella Setya Kartika¹, Yuliant Sibaroni², Adiwijaya³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹riemameristella@student.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id,

³adiwijaya@telkomuniversity.ac.id

Abstrak

Dengan berkembangnya teknologi menyebabkan *e-commerce* semakin berkembang pesat. *Startup e-commerce* membangun toko online yang menjual berbagai macam kebutuhan seperti kosmetik, makanan, pakaian dan lain-lain. Ulasan produk menjadi salah satu hal yang perlu dipertimbangkan sebelum membeli suatu produk. Ulasan produk pun beraneka ragam, ada yang cenderung ke positif atau negatif. Ulasan produk berjumlah ratusan bahkan ribuan. Maka dari itu dibutuhkan waktu yang lama untuk membaca ulasan produk satu persatu secara manual. Untuk menyelesaikan masalah tersebut diusulkan klasifikasi sentimen untuk mengklasifikasikan ulasan produk kedalam kategori positif atau negatif menggunakan metode *Multinomial Naïve Bayes* yang efisien dalam komputasi. Untuk menghasilkan F1-Score terbaik, perbandingan tahap *preprocessing* dilakukan antara *preprocessing* menggunakan stemming, *preprocessing* menggunakan lemmatisasi dan *preprocessing* tanpa stemming dan lemmatisasi, setelah itu dilakukan perbandingan pada tahap seleksi fitur menggunakan *Information Gain* dan *Query Expansion Ranking(QER)*. Hasil percobaan terbaik adalah model dengan *preprocessing* tanpa stemming dan lemmatisasi serta tanpa menggunakan seleksi fitur dan model dengan *preprocessing* tanpa stemming dan lemmatisasi serta menggunakan seleksi fitur *Information Gain* pada penggunaan 91% fitur dengan rata-rata F1-Score yang didapatkan adalah 93,38%.

Kata kunci : klasifikasi sentimen, *Multinomial Naïve Bayes*, stemming, lemmatisasi, *Information Gain*, *Query Expansion Ranking*.

Abstract

The development of technology led to *e-commerce* growing rapidly. *Startup e-commerce* builds an online store that sells a variety of needs such as cosmetics, food, clothing and etc. Product reviews become one of the things to consider before buying a product. Product reviews are diverse, it can be positive or negative. The product reviews can be hundreds or even thousands of reviews. Therefore it takes a long time to read product reviews one by one manually. To solve this problem, a sentiment classification is proposed to classification product reviews into positive or negative categories using the *Multinomial Naïve Bayes* method which efficient in computing. To produce the best F1-Score, comparison of preprocessing stage was tried between preprocessing using stemming, preprocessing using lemmatization and preprocessing without stemming and lemmatization, then a comparison was made on the feature selection stage using *Information Gain* and *Query Expansion Ranking(QER)*. The best results from these experiments are model with preprocessing without stemming and lemmatization also without feature selection and model with preprocessing without stemming and lemmatization also using *Information Gain* feature selection on the use of 91% features with an average F1-Score obtained is 93.38%

Keywords: sentiment classification, *Multinomial Naïve Bayes*, stemming, lemmatization, *Information Gain*, *Query Expansion Ranking*.
