

Feature Selection Menggunakan Chi-Square Dan N-Gram Dengan Algoritma Naïve Bayes Classifier Untuk Analisis Sentimen Review Produk Elektronik

¹*Deni Irvantoro (1410651199)*

²*Ilham Saifuddin S.Pd., M.Si.*

³*Reni Umilasari S.Pd., M.Si.*

*Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
Email : den.irvanto99@gmail.com*

ABSTRAK

Analisis sentimen perlu diterapkan untuk menentukan review yang bernilai positif atau negatif. Algoritma klasifikasi seperti Naïve Bayes Classifier (NBC) telah banyak di terapkan. Di samping memiliki performanya yang baik di berbagai domain, tapi terdapat sensitifitas terhadap pemilihan fitur. Chi-square merupakan salah satu metode statistik untuk pemilihan fitur dengan uji independensi dan estimasi yang bertujuan untuk mengetahui ketergantungan suatu class pada suatu fitur. Sedangkan N-Gram merupakan pemotongan kata berdasarkan nilai n. Dengan penggabungan chi-square dan N-Gram pada teknik uni-gram hasil tertinggi terdapat pada uji 75% fitur dengan akurasi 89%, presisi 89%, dan recall 100%. Hasil tertinggi pada bi-gram terdapat pada 25% fitur dengan akurasi 89%, presisi 89%, dan recall 100%. Hasil tertinggi pada uni+bi-gram terdapat pada 25% fitur dengan akurasi 89%, presisi 89%, dan recall 100%. Dari ketiga teknik N-Gram improvisasi terjadi pada fitur uji 25% dengan akurasi uni-gram 87,6% menjadi 89% pada bi-gram dan uni+bi-gram, presisi uni-gram 88,84% menjadi 89% pada uni-gram+ bi-gram, dan recall uni-gram 98,42% menjadi 100% pada bi-gram dan uni+bi-gram. Dalam penelitian ini masing-masing teknik N-Gram dengan penerapan feature selection chi-squre memiliki ambang batas penggunaan fitur yang berbeda. Pada komparasi ke tiga teknik N-Gram secara umum chi-square tidak mempengaruhi urutan hasil dari klasifikasi, tapi mampu memberikan improvisasi hasil akurasi, presisi, dan recall.

Kata Kunci: Analisis Sentimen, Feature Selection, Chi-Square, N-Gram, Agoritma Naïve Bayes Classifier..

Feature Selection Using Chi-Square and N-Gram with Naïve Bayes Classifier Algorithm for Sentiment Analysis of Electronic Product Reviews

¹*Deni Irvantoro (1410651199)*

²*Ilham Saifuddin S.Pd., M.Si.*

³*Reni Umilasari S.Pd., M.Si.*

*Informatics Engineering Department, Faculty of Engineering, University of
Muhammadiyah Jember
Email : den.irvanto99@gmail.com*

ABSTRACT

Sentiment analysis needs to be applied to determine reviews that are positive or negative. Classification algorithms such as Naïve Bayes Classifier (NBC) have been applied. In addition, it has good performance in various domains, but there is sensitivity to feature selection. Chi-square is one of the statistical methods for feature selection with an independence and estimation test that aims to determine the dependence of a class on a feature. While the N-Gram is a word cut based on the value of n. With the incorporation of chi-square and N-Gram on the uni-gram technique the highest results are found in the 75% test feature with 89% accuracy, 89% precision, and 100% recall. The highest results on bi-gram are in 25% features with 89% accuracy, 89% precision, and 100% recall. The highest results in uni + bi-gram are found in 25% features with 89% accuracy, 89% precision, and 100% recall. Of the three improvisation techniques N-Gram occurred in the test feature 25% with uni-gram accuracy 87.6% to bi-gram and uni + bi-gram, uni-gram precision 88.84% to 89% in uni-gram + bi-gram, and uni-gram recall 98.42% to 100% in bi-gram and uni+bi-gram. In this study each N-Gram technique with the application of the chi-square selection feature has a different feature usage threshold. In comparison to the three N-Gram techniques in general chi-square does not affect the order of results from the classification, but is able to provide improvisation results of accuracy, precision, and recall.

Keywords: *Sentiment Analysis, Feature Selection, Chi-Square, N-Gram, Naïve Bayes Classifier Algorithm.*