

INTISARI

ANALISIS SENTIMEN MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN SELEKSI FITUR *CHI SQUARE* DENGAN PENANGANAN DATA *IMBALANCED* : STUDI KASUS *REVIEW PRODUK SKINCARE* PADA *PLATFORM ONLINE SOCIOLLA*

Oleh

Amaliya Nur Mahmudah

18/424278/PA/18383

Ulasan produk sangat berpengaruh terhadap citra perusahaan dan keputusan pembelian konsumen terhadap suatu produk *skincare*. Banyaknya pengguna dan ulasan produk, sehingga sulit untuk diproses secara manual. Oleh sebab itu, pada penelitian ini dilakukan analisis sentimen untuk mengklasifikasikan ulasan ke dalam kelas sentimen positif dan negatif secara otomatis. Data ulasan yang berbentuk teks menyebabkan data memiliki dimensi yang tinggi dan fitur yang sangat banyak. Pemilihan fitur yang baik dapat meningkatkan performa klasifikasi serta mampu mempersingkat waktu proses klasifikasi.

Pada penelitian ini akan dibandingkan performa klasifikasi *Naïve Bayes* pada data *imbalanced* dan performa *Naïve Bayes* seleksi fitur *Chi Square* dengan penanganan data *imbalanced*. Masalah *imbalanced* data dapat menyebabkan hasil dari klasifikasi yang tidak akurat atau biasa disebut dengan *accuracy paradox*. Untuk mengatasi hal tersebut diperlukan suatu metode *resampling* yaitu teknik *random oversampling*.

Kata Kunci : analisis sentimen, *Naïve Bayes*, *Chi Square*, data *imbalanced*, *random oversampling*, seleksi fitur.

ABSTRACT

SENTIMENT ANALYSIS USING NAÏVE BAYES ALGORITHM AND CHI SQUARE FEATURE SELECTION WITH IMBALANCED DATA HANDLING : A CASE STUDY OF SKINCARE PRODUCT REVIEWS ON THE SOCIOLLA ONLINE PLATFORM

By

Amaliya Nur Mahmudah

18/424278/PA/18383

Product reviews are very influential on company image and consumer purchasing decisions for a skincare product. There are many users and product reviews, making it difficult to process manually. Therefore, in this research, sentiment analysis is carried out to classify reviews into positive and negative sentiment classes automatically. Review data in the form of text causes the data to have high dimensions and a lot of features. Good feature selection can improve classification performance and can shorten the classification process time.

This research will compare the performance of Naïve Bayes classification on imbalanced data and the performance of Naïve Bayes Chi Square feature selection with imbalanced data handling. Imbalanced data problems can cause inaccurate classification results or what is commonly called the accuracy paradox, it is necessary to approach by resampling on the original data using the imbalanced oversampling techniques.

Keywords : sentiment analysis, Naïve Bayes, Chi Square, imbalanced data, random oversampling, feature selection.