

INTISARI

IMPLEMENTASI *NAÏVE BAYES* DAN *RANDOM FOREST* UNTUK ANALISIS SENTIMEN TERHADAP DATA *IMBALANCED REVIEW* PRODUK KOSMETIK PADA *PLATFORM ONLINE SOCIOLLA*

Oleh

Lely Nur Aisah
16/398660/PA/17621

Perkembangan teknologi yang terus meningkat dalam era *Big Data* ini menjadi salah satu perintis munculnya situs jual beli online di Indonesia. Salah satu situs jual beli online yang terkenal di kalangan anak muda adalah *Sociolla*, situs jual beli produk kosmetik terlengkap dan terpercaya. Tidak hanya membeli produk kosmetik, pada situs tersebut konsumen juga dapat memberikan ulasan terhadap produk yang mereka beli. Tidak bisa dipungkiri bahwa ulasan yang muncul dapat mempengaruhi citra dari perusahaan. Penting bagi perusahaan untuk mengetahui tanggapan publik mengenai produk yang mereka tawarkan. Tanggapan publik tersebut berjumlah terlalu banyak untuk diproses secara manual. Oleh sebab itu, diperlukan metode yang mampu mengklasifikasikan ulasan ke dalam sentimen positif dan negatif secara otomatis. Metode klasifikasi yang dapat digunakan adalah *Naïve Bayes* dan *Random Forest*. Dalam proses klasifikasi data ulasan secara *real*, tentunya akan menemui data *imbalanced*, dimana jumlah data pada kelas sentimen positif berjumlah lebih banyak dibandingkan dengan jumlah data pada kelas sentimen negatif. Maka diperlukan metode yang dapat mengatasi hal tersebut, yaitu dengan teknik *undersampling*. Dari perbandingan nilai akurasi, sensitivitas, dan spesifisitas diperoleh kesimpulan bahwa metode *Random Forest Undersampling* merupakan metode terbaik untuk klasifikasi pada kasus ini.

Kata kunci : Analisis sentimen, *review*, *web scraping*, data *imbalanced*, *Naïve Bayes*, *Random Forest*, teknik *undersampling*

ABSTRACT

IMPLEMENTATION OF NAÏVE BAYES AND RANDOM FOREST FOR SENTIMENT ANALYSIS OF IMBALANCED DATA COSMETIC PRODUCT REVIEWS ON THE SOCIOLLA ONLINE PLATFORM

By

Lely Nur Aisah

16/398660/PA/17621

The growth of technology in the Big Data era has become one of the pioneers of the rise of e-commerce in Indonesia. One of the most popular e-commerce platforms among young people is Sociolla, which is the most complete and trustworthy e-commerce for buying cosmetic products. On this site, aside from buying products, consumers can also write reviews of product they have bought. Review can affect company's image in public eye. It is important for company to know about how public responds about their product. The amount of reviews is too much to be processed manually. Therefore, a special method is needed to classify the reviews automatically, whether it is poitive or negative. We can analyze it using the Naïve Bayes and Random Forest method to classify review data. However, because the dataset that is being used is an imbalanced data, where the amount of positive class is greater than the amount of negative class, it is necessary to approach by resampling on the original data using the imbalanced undersampling techniques. After using the undersampling technique, we can conclude that Random Forest method is the best method for this case.

Keyword : sentiment analysis, review, web scraping, imbalanced data, Naïve Bayes, Random Forest, undersampling technique