

ABSTRAK

Berkembangnya industri kecantikan membuat produk *skincare* semakin beragam, toner dengan bahan utama *Centella Asiatica* menjadi salah satu produk *skincare* yang mendapatkan perhatian khusus. Semakin beragamnya produk toner yang tersedia di pasar *skincare*, semakin meningkat pula selektivitas konsumen dalam pemilihan produk. *Female Daily* merupakan *platform* yang digunakan untuk menampung segala *review* produk *skincare*. Namun, jumlah *review* yang tersedia sangat banyak, membuat konsumen membutuhkan waktu yang cukup lama dalam membaca secara keseluruhan *review*.

Penelitian ini dapat membantu konsumen sebagai bahan pertimbangan pembelian suatu produk melalui Analisis sentimen dengan mengimplementasikan dan menganalisa performa metode *Latent Dirichlet Allocation* untuk pemodelan topik dan metode *Support Vector Machine* untuk mengklasifikasikan sentimen data ulasan produk *skincare* ke dalam dua kelas yaitu positif dan negatif. Tahapan analisis sentimen pada penelitian ini dimulai dengan pengumpulan data dengan *web scraping* ulasan pada platform Female Daily dengan jumlah 1000 ulasan. Selanjutnya, *preprocessing* untuk menghasilkan data yang terstruktur. Pada pemodelan topik menggunakan metode *Latent Dirichlet Allocation* dan labelisasi ulasan menggunakan *InSet Lexicon*. Kemudian, dilakukan pembobotan kata menggunakan TF-IDF dan data *splitting* sebelum data diseimbangkan pada setiap kelasnya menggunakan SMOTE, data dilakukan proses pembagian data latih dan data uji dengan perbandingan 80:20. Selanjutnya dilakukan pemodelan klasifikasi menggunakan *support vector machine*, kemudian dilakukan pengujian menggunakan *confusion matrix* dan *k-fold cross validation* untuk mengevaluasi kinerja model klasifikasi, mengetahui tingkat akurasi, dan keberhasilan model klasifikasi.

Penelitian ini menghasilkan tiga aspek yaitu Penggunaan Produk, Kualitas Produk, dan Pembelian Produk. LDA menghasilkan *coherence score* tertinggi 0.333 berdasarkan distribusi jumlah topik pada *num topic* 3 dari 2-10. Kemudian, labeling *InSet Lexicon* menghasilkan 392 ulasan positif, dan 608 ulasan negatif. Hasil pengujian menggunakan *confusion matrix* dan *k-fold cross validation*, algoritma SVM memperoleh akurasi sebesar 82,5%, rata-rata presisi sebesar 82,03%, dan rata-rata recall sebesar 80,29%, sementara pada pengujian *k-fold* dengan K=5 memperoleh akurasi tertinggi sebesar 80.63% pada *fold* ke dua dan ke tiga. Kemudian, dilakukan resampling dengan hasil akurasi 84,5%, rata-rata presisi sebesar 83,51%, dan rata-rata recall sebesar 83,68%. Pada pengujian *k-fold* dengan K=5 menggunakan SMOTE menghasilkan akurasi *k-fold* tertinggi sebesar 89,64 % pada *fold* ke lima.

Kata kunci : Analisis Sentimen Berbasis Aspek, Skincare, Latent Dirichlet Allocation, Support Vector Machine, SMOTE

ABSTRACT

The development of the beauty industry makes skincare products more diverse, toner with Centella Asiatica as the main ingredient is one of the skincare products that get special attention. The more diverse toner products available in the skincare market, the more selectivity consumers have in choosing products. Female Daily is a platform used to accommodate all skincare product reviews. However, the number of reviews available is very large, making consumers take a long time to read the entire review.

This research can help consumers as a consideration for purchasing a product through sentiment analysis by implementing and analyzing the performance of the Latent Dirichlet Allocation method for topic modeling and the Support Vector Machine method for classifying the sentiment of skincare product review data into two classes, namely positive and negative. The sentiment analysis stages in this study began with data collection by web scraping reviews on the Female Daily platform with a total of 1000 reviews. Next, preprocessing to produce structured data. Topic modeling uses the Latent Dirichlet Allocation method and review labeling using InSet Lexicon. Then, word weighting is done using TF-IDF and data splitting before the data is balanced in each class using SMOTE, the data is divided into training data and test data with a ratio of 80:20. Next, classification modeling is carried out using support vector machine, then testing is carried out using confusion matrix and k-fold cross validation to evaluate the performance of the classification model, determine the level of accuracy, and the success of the classification model.

This research resulted in three aspects, namely Product Usage, Product Quality, and Product Purchase. LDA produces the highest coherence score of 0.333 based on the distribution of the number of topics on num topic 3 from 2-10. Then, InSet Lexicon labeling produced 392 positive reviews, and 608 negative reviews. The test results using confusion matrix and k-fold cross validation, the SVM algorithm obtained an accuracy of 82.5%, an average precision of 82.03%, and an average recall of 80.29%, while in k-fold testing with $K = 5$ obtained the highest accuracy of 80.63% in the second and third folds. Then, resampling is done with an accuracy of 84.5%, an average precision of 83.51%, and an average recall of 83.68%. In k-fold testing with $K = 5$ using SMOTE, the highest k-fold accuracy was 89.64% in the fifth fold.

Keywords: *Aspect-based Sentiment Analysis, Skincare, Latent Dirichlet Allocation, Support Vector Machine, SMOTE*