

ABSTRAK

Bisnis *coffee shop* yang berkembang dengan sangat pesat menyebabkan terjadinya persaingan antar *coffee shop*, untuk mengimbangi persaingan tersebut *coffee shop* akan berusaha memberikan kualitas dan pelayanan terbaik dengan memanfaatkan ulasan dari pelanggan yang menjadi tolak ukur utama dalam meningkatkan kualitas produk dan layanan. Ulasan pelanggan merupakan hal penting yang berdampak besar pada citra sebuah bisnis, sayangnya ada banyak ulasan yang bersifat ambigu, pada *coffee shop* yang memiliki banyak maka cabang klasifikasi secara manual akan sangat sulit dilakukan.

Untuk mengatasi permasalahan tersebut, penelitian ini melakukan analisis sentimen terkait ulasan pelanggan dengan mengkombinasikan metode *lexicon based* untuk pelabelan data dan metode *support vector machine* sebagai model klasifikasi yang dilatih menggunakan label dari metode *lexicon-based*. Pada penelitian ini analisis sentimen dilakukan pada ulasan pelanggan *coffee shop* Couve, dengan melakukan teknik *web scraping* didapatkan dataset sejumlah 2.471 ulasan.

Data kemudian diproses melalui beberapa tahap *preprocessing* seperti *case folding*, *data cleansing*, *tokenizing*, *stemming*, dan penghapusan *stopword*. Pelabelan data dilakukan dengan menggunakan pendekatan berbasis leksikon dari *VADER*. Metode SVM diterapkan untuk melakukan klasifikasi sentimen yang dibagi menjadi tiga kelas: positif, netral, dan negatif. Hasil pengujian model dengan *Confusion Matrix* pada metode SVM-VADER mencapai akurasi sebesar 82%, presisi 78%, dan *recall* 95%. Setelah menerapkan teknik *SMOTE* untuk menangani masalah ketidakseimbangan data, hasil akhir menunjukkan peningkatan akurasi menjadi 94%, presisi 91%, dan *recall* 93%. Hal ini menunjukkan bahwa penerapan metode SVM dengan pendekatan *Lexicon Based* dan *SMOTE* mampu meningkatkan performa analisis sentimen secara signifikan

Kata kunci: *Support Vector Machine, Lexicon Based, SMOTE, VADER*

ABSTRACT

The rapid growth of the coffee shop business has led to increasing competition among coffee shops. To keep up with this competition, coffee shops strive to provide the best quality and service by leveraging customer reviews, which serve as a key benchmark for improving product and service quality. Customer reviews are crucial, as they significantly impact a business's reputation. Unfortunately, many reviews are often ambiguous, making manual classification particularly challenging for coffee shops with multiple branches.

To solve those problems, this research conducts sentiment analysis on customer reviews by combining a lexicon-based method for data labeling with a support vector machine (SVM) model trained using labels generated by the lexicon-based approach. The sentiment analysis in this study focuses on customer reviews of the Couve coffee shop. By utilizing web scraping techniques, a dataset of 2,471 reviews was obtained.

The data underwent several preprocessing steps, including case folding, data cleansing, tokenizing, stemming, and stopword removal. The data labeling was performed using the lexicon-based VADER approach. The SVM method was applied to classify the sentiments into three categories: positive, neutral, and negative. The model evaluation using a confusion matrix on the SVM-VADER method achieved an accuracy of 82%, a precision of 78%, and a recall of 95%. After applying the SMOTE technique to address data imbalance, the final results showed an increase in accuracy to 94%, precision to 91%, and recall to 93%. This demonstrates that the application of the SVM method with the Lexicon-Based and SMOTE approaches significantly improves sentiment analysis performance

Keywords: Support Vector Machine, Lexicon Based, SMOTE, VADER