

INTISARI

Ulasan *customer* pada sebuah produk yang merupakan bagian dari identifikasi *customer needs* sangat berpengaruh dalam proses pengembangan produk. Namun, metode identifikasi kebutuhan pelanggan secara konvensional, terlebih dalam tahap pengumpulan *voice of customer*, belum memberikan performa yang baik dalam efisiensi waktu dan biaya yang digunakan. Guna meningkatkan performa tahapan tersebut, banyak dikembangkan metode analisis sentimen untuk memahami *feedback* konsumen pada ulasan *online* produk, yang saat ini memiliki jumlah yang sangat banyak akibat maraknya perkembangan teknologi. Maka dari itu, tidak sedikit *customer* memberikan kritik atau pendapatnya melalui media internet. Pada penelitian sebelumnya, ditemukan bahwa kerangka kerja yang digunakan mampu memberikan informasi terkait tendensi sentimen *customer* pada fitur produk tertentu, akan tetapi dalam penentuan fitur produk tidak dilakukan secara *unsupervised* sehingga besar kemungkinan tidak semua fitur dalam ulasan teridentifikasi. Tujuan dari penelitian ini adalah untuk mengidentifikasi fitur suatu produk dan menganalisis sentimen pada suatu ulasan *online* secara *unsupervised* yang memiliki karakteristik mampu mengidentifikasi fitur yang belum ditangkap oleh penulis serta memberikan visualisasi hasil berupa kelompok fitur dengan pasangan kata dari masing-masing kelompok fitur tersebut. Pasangan kata ini bertujuan untuk mengetahui penyebab penilaian oleh *customer* terhadap suatu fitur produk. Pengelompokan fitur produk dalam penelitian ini menggunakan metode *Hierarchical Clustering*.

Dalam penelitian ini berfokus kepada perancangan kerangka kerja identifikasi sebuah fitur dan analisis sentimen pada ulasan *online* sebuah produk dengan menggunakan bahasa pemrograman Python. Proses identifikasi dan analisis sentimen ini dilakukan dengan metode *unsupervised*. Adapun terdapat tujuh tahapan utama yang harus dilakukan, yaitu: 1) Tahap Pengambilan Data, 2) Tahap Seleksi Data, 3) Tahap Pemrosesan Data Awal, 4) Tahap Analisis Data, 5) Tahap Klasterisasi Data, 6) Tahap Validasi dan 7) Tahap Visualisasi Hasil. Guna melakukan seluruh tahapan tersebut, terdapat dua domain utama yang digunakan, yaitu 1) Domain aktif, yang terdiri atas analisis dan *designer*, dan 2) Domain pasif, yang terdiri atas *platform e-commerce* dan NLP.

Hasil penelitian ini menunjukkan performa analisis sentimen yang digunakan pada penelitian ini dengan menggunakan algoritma TextBlob, memiliki nilai akurasi sebesar 94,40%, presisi sebesar 97,50%, *recall* sebesar 96,27%, dan *F1-score* sebesar 96,88%.

Kata kunci: *Hierarchical Clustering, Identification Customer Needs, Natural Language Processing, Online Review, Product Feature Identification, Sentiment Analysis, Unsupervised Learning, Voice of Customer, Word Similarity*

ABSTRACT

Customer reviews of a product that are part of the identification of customer needs are very influential in the product development process. However, conventional methods of identifying customer needs, especially in the stage of gathering voice of customers, have not provided good performance in terms of time and cost efficiency. Then, for increasing the performance of existing method, sentiment analysis has been developed much to understand customer feedback in online product reviews, which currently have an enormous amount due to the proliferation of technological developments. Therefore, not a few customers give criticism or opinions through the internet media. In previous studies, it was found that the framework used is able to provide information related to customer sentiment tendencies on certain product features, but in determining product features not done unsupervised so it is probable that not all features in the review are identified. The purpose of this study is to identify the features of a product and analyze sentiments on an unsupervised online review that has the characteristics of being able to identify features that have not been captured by the author and provide a visualization of the results in the form of a group of features with pairs of words from each of those feature groups. This pair of words aims to find out the cause of customer assessment of a product's features. Grouping of product features in this research uses the Hierarchical Clustering method.

This research focuses on designing a feature identification framework and sentiment analysis on online product reviews using the Python programming language. The process of identifying and analyzing sentiments is done by the unsupervised method. There are seven main stages that must be carried out, namely: 1) Data Retrieval Phase, 2) Data Selection Phase, 3) Initial Data Processing Phase, 4) Data Analysis Phase, 5) Data Clustering Phase, 6) Validation Phase and 7) Data Visualization Phase. To carry out all these stages, there are two main domains used, namely 1) Active domain, which consists of analysts and designers, and 2) Passive domain, which consist of platforms e-commerce and NLP.

The results of this study show the sentiment analysis performance used in this study using the TextBlob algorithm, has an accuracy value of 94.40%, precision of 97.50%, recall of 96.27%, and F1-score of 96.88%.

Keyword: Hierarchical Clustering, Identification Customer Needs, *Natural Language Processing*, Online Review, Product Feature Identification, Sentiment Analysis, Unsupervised Learning, Voice of Customer, Word Similarity