

## INTISARI

Pengembangan model mobil baru sebagai strategi untuk mempertahankan daya saing dalam pasar yang kompetitif menghadapi tantangan karena dinamika pasar dan siklus perubahan model mobil yang semakin cepat. Akibat tantangan tersebut, pengumpulan informasi kebutuhan konsumen untuk mendukung kesuksesan pengembangan mobil baru tidak lagi efektif dilakukan secara tradisional, sehingga diperlukan pemanfaatan sumber informasi alternatif, seperti data dari media sosial. Pemodelan topik menjadi pendekatan populer untuk identifikasi informasi penting dari data besar media sosial, tetapi interpretasi topik masih menjadi kendala, terutama saat topik diperlukan untuk analisis lebih lanjut. Penelitian ini bertujuan untuk menyediakan alat identifikasi dan analisis suara konsumen dari data besar di media sosial menjadi peluang ide peningkatan atribut mobil generasi terbaru.

Metode hibrida yang mengintegrasikan pelabelan topik tanpa pengawasan dan pemodelan peluang diperkenalkan untuk menganalisis komentar pengguna media sosial. Prosedur pelabelan topik berbasis pemodelan topik dan aturan kombinasi kelas kata menghasilkan label yang mudah dipahami serta mampu menangkap ekspresi emosional sehingga dapat secara langsung dianalisis menggunakan pemodelan peluang untuk menghasilkan ide-ide bagi peningkatan desain atribut produk otomotif. Tujuh algoritma pemodelan topik, seperti LSI, NMF, LDA, BTM, modifikasi Word2Vec, modifikasi BERT, *ClusteringVector*, dan tiga aturan kombinasi kelas kata, seperti penggunaan kata benda, kombinasi kata benda dan kata sifat, serta kombinasi kata benda, kata sifat, dan kata kerja dievaluasi untuk mendapatkan prosedur pelabelan topik terbaik. Prosedur terbaik ditetapkan berdasarkan tingkat kemiripan label topik yang dihasilkan terhadap topik yang ditetapkan secara manual oleh 6 evaluator menggunakan data sampel. Eksperimen dilakukan untuk menunjukkan efektivitas prosedur terbaik menggunakan data komentar pengguna YouTube pada konten video ulasan mobil Toyota Avanza generasi kedua.

Pendekatan *Non-Negative Matrix Factorization* dengan kombinasi kata benda dan kata sifat ditetapkan sebagai prosedur pelabelan topik otomatis paling efektif untuk menghasilkan label topik yang menyerupai topik oleh evaluator dengan tingkat kemiripan 0,43 dan 0,13 untuk *cosine* dan *jaccard similarity*. Prosedur terbaik digunakan untuk menganalisis 5.131 data komentar sehingga menghasilkan 21 label topik yang sesuai dan sembilan topik diantaranya secara eksplisit terkait dengan atribut mobil. Pemetaan peluang menunjukkan enam topik berada di kuadran *underserved*, satu topik di kuadran *served right*, dan dua topik di kuadran *overserved*, yang menunjukkan 83,3% topik selaras dengan strategi perubahan minor model mobil terbaru. Hasil penelitian memperkuat studi terdahulu terkait pemanfaatan data media sosial sebagai representasi suara konsumen dan berkontribusi terhadap kinerja aktivitas peningkatan desain atribut kendaraan, sekaligus meningkatkan efisiensi keseluruhan proses desain.

**Keywords:** klasifikasi sentimen; opportunity algorithm; pelabelan topik otomatis; pemodelan topik; peningkatan desain produk.

## **ABSTRACT**

*The development of new car models as a strategy to maintain competitiveness in an increasingly dynamic market faces significant challenges due to rapid market changes and shorter model life cycles. Consequently, the traditional approaches to collecting consumer requirement information to support the success of new car development have become less effective. This situation necessitates the utilization of alternative information sources, such as data derived from social media. Topic modeling has emerged as a popular approach for identifying key information within large-scale social media data; however, topic interpretation remains a major challenge, particularly when topics are required for further analysis. This study aims to provide a framework for identifying and analyzing consumer voices extracted from social media big data, transforming them into potential ideas for enhancing the attributes of next-generation automobiles.*

*A hybrid method integrating unsupervised topic labeling and opportunity modeling was introduced to analyze social media user comments. The topic labeling procedure, based on topic modeling and part-of-speech combination rules, produces interpretable labels that effectively capture emotional expressions. These labels can then be directly analyzed through opportunity modeling to generate insights for improving the design of automotive product attributes. Seven topic modeling algorithms—namely LSI, NMF, LDA, BTM, modified Word2Vec, modified BERT, and ClusteringVector—and three part-of-speech combination rules—namely the use of nouns, the combination of nouns and adjectives, and the combination of nouns, adjectives, and verbs—were evaluated to identify the most effective topic labeling procedure. The optimal procedure was determined based on the degree of similarity between the automatically generated topic labels and manually assigned topics by six evaluators using sample data. Experiments were conducted to demonstrate the effectiveness of the selected procedure using YouTube user comments on video reviews of the second-generation Toyota Avanza.*

*The Non-Negative Matrix Factorization approach combined with noun–adjective pair extraction was identified as the most effective automatic topic labeling procedure, producing topic labels most similar to those identified by the evaluators, with similarity scores of 0,43 (cosine similarity) and 0,13 (Jaccard similarity). The optimal procedure was applied to analyze 5.131 comment data, resulting in 21 coherent topic labels, nine of which were explicitly related to automotive attributes. The opportunity mapping indicated that six topics were positioned in the underserved quadrant, one topic in the served right quadrant, and two topics in the overserved quadrant, demonstrating that 83,3% of the topics were aligned with the minor model change strategy of the latest vehicle. These findings reinforce prior studies on the use of social media data as a representation of the consumer voice and contribute to enhancing the performance of vehicle attribute design improvement activities, while simultaneously increasing the overall efficiency of the design process.*

**Keywords:** *automatic topic labeling; opportunity algorithm; product design improvement; sentiment classification; topic modeling.*