

## ABSTRACT

### MULTIMODAL DEEP LEARNING-BASED PRODUCT MATCHING MODEL USING EFFICIENTNET & ROBERTA ENCODERS WITH LATE FUSION TECHNIQUE

Oleh

Achmad Hadzami Setiawan

21/480222/PA/20851

Product matching in e-commerce poses a challenge in identifying identical products that may have variations in naming, descriptions, or images. Text-based or image-based approaches separately often face limitations in capturing the semantic relationships between products. Therefore, this study aims to develop a multimodal product matching model that integrates visual and textual information using EfficientNet, RoBERTa, and K-Nearest Neighbors (KNN).

This study proposes the use of EfficientNet to extract visual features from product images, while RoBERTa is employed to extract features from product titles. The representations from both modalities will be combined using the late fusion strategy, with three experimental schemes to be tested. Subsequently, the extracted features will be used in KNN-based classification utilizing Cosine Similarity as the distance metric. Additionally, this study implements a supervised contrastive learning approach to enhance embedding quality before classification.

Experiments will be conducted on a Shopee e-commerce product dataset to evaluate the effectiveness of the proposed model. The evaluation will be based on precision, recall, and F1-score metrics to measure how well the model matches similar products. The findings of this study are expected to contribute to improving the performance of product matching systems in e-commerce through a multimodal approach.

**Keywords:** *Product Matching, Multimodal, EfficientNet, RoBERTa, Late fusion, K-Nearest Neighbours, Deep Learning, E-commerce.*

## INTISARI

### MODEL PENCOCOKAN PRODUK BERBASIS PEMBELAJARAN MESIN MENDALAM MULTIMODAL MENGGUNAKAN *ENCODER* EFFICIENTNET & ROBERTA DENGAN TEKNIK *LATE FUSION*

Oleh

Achmad Hadzami Setiawan

21/480222/PA/20851

Pencocokan produk dalam *e-commerce* merupakan tantangan dalam mengidentifikasi produk yang sama namun memiliki variasi dalam penamaan, deskripsi, atau gambar. Pendekatan berbasis teks atau gambar secara terpisah sering kali mengalami keterbatasan dalam menangkap hubungan semantik antar produk. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model pencocokan produk berbasis multimodal yang menggabungkan informasi visual dan tekstual menggunakan EfficientNet, RoBERTa, dan *K-Nearest Neighbors* (KNN).

Penelitian ini mengusulkan penggunaan EfficientNet untuk mengekstrak fitur visual dari gambar produk, sementara RoBERTa digunakan untuk mengekstrak fitur dari judul produk. Representasi dari kedua modalitas akan digabungkan menggunakan strategi *late fusion*, dengan tiga skema eksperimen yang akan diuji. Setelah itu, fitur yang diperoleh akan digunakan dalam klasifikasi berbasis KNN menggunakan metrik jarak kesamaan kosinus. Penelitian ini juga menerapkan pendekatan pembelajaran kontrasif dengan supervisi untuk meningkatkan kualitas *embedding* sebelum klasifikasi dilakukan.

Eksperimen akan dilakukan pada dataset produk *e-commerce* Shopee untuk mengevaluasi efektivitas model yang diusulkan. Evaluasi dilakukan berdasarkan metrik presisi, *recall*, dan skor F1 untuk mengukur seberapa baik model dapat mencocokkan produk yang serupa. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan performa sistem pencocokan produk di *e-commerce* melalui pendekatan multimodal.

**Kata Kunci:** *Pencocokan Produk, Multimodal, EfficientNet, RoBERTa, Late fusion, K-Nearest Neighbour, Pembelajaran Mesin Mendalam, E-commerce*