Multi-Modal Product Matching with Deep Neural Networks for Large-Scale E-Commerce Applications

WHY THIS PROJECT

- Duplicate product listings confuse buyers, mislead pricing, and reduce platform trust.
- Sellers are affected by copycat listings and price manipulation.
- Visual and textual differences make traditional matching methods ineffective.
- Our system combines image and text embeddings using deep learning to detect duplicates.
- This solution helps improve search relevance, reduce fraud, and clean up product catalogs at scale.

TEAM

Kashish Vatsal Desai
Khushi Nilanshu Patel
Shreya Pudukkottai
Sravani Neelapala
Ameya Panchpor
Jeevan Shankar Balagam

OBJECTIVE

- Understand multimodal challenges in product matching
- Develop & compare multiple deep learning models
- Evaluate performance using F1 Score, Precision, Recall
- Ensure scalability & real-world applicability

MODELS

Model 1: *ECA NFNet L1 + Paraphrase-XLM-R + FAISS + INB*
*I*mage Model: ECA NFNet L1 (Efficient Channel Attention + Normalizer-Free ResNet)
Model 2 : *NFNet + Swin Transformer + EfficientNet + Distil-BERT + ALBERT + Multilingual BERT + KNN Voting*
Image Models: NFNet, Swin Transformer, EfficientNet
Text Models: Distil-BERT, ALBERT, Multilingual BERT, TF-IDF
Model 3: *ViT + NFNet-F0 + Indonesian-BERT + Multilingual-BERT + Paraphrase-XLM + GAT (planned)*
Image Models: ViT (Vision Transformer), NFNet-F0
Text Models: Indonesian-BERT, Multilingual-BERT, Paraphrase-XLM

Model	F1 Score	Precision	Recall
Model 1	**0.9097**	0.8657	0.8731
Model 2	0.14	0.554	0.8135
Model 3	0.5324	0.4271	0.8858

DATASET

- Source: Shopee Product Matching Kaggle Dataset
- 34,251 product listings, 32,412 images
- Features: posting_id, image, title, label_group
- Challenges: multilingual noise, class imbalance, scalability