Product Price Prediction — Final Report

Overview

This project presents a multimodal machine learning solution for predicting product prices by analyzing both textual (catalog content) and visual (product images) data. It combines BERT embeddings, CNN-based visual features, and ensemble models to achieve high predictive accuracy under SMAPE evaluation.

Dataset Summary

- Training Samples: 75,000
- Test Samples: 75,000
- Columns: sample_id, catalog_content, image_link, price
- Objective: Predict product price (price) for unseen items
- Distribution: Log-normal (skewed toward low-price items)

Methodology

1. Exploratory Data Analysis (EDA): Identified log-normal price distribution and structured text patterns.
2. Text Processing (NLP Pipeline): Used BERT embeddings and TF-IDF for text representation. Extracted brand, specifications, weight, and category indicators.
3. Image Processing (CV Pipeline): Applied ResNet18 for 512-dimensional image feature extraction.
4. Multimodal Fusion: Combined text, visual, and engineered features with RobustScaler normalization.
5. Modeling and Ensembles: XGBoost, LightGBM, Random Forest, ElasticNet, Neural Network, and Stacking Ensemble.

Implementation Details

Component	Tools / Libraries Used
NLP	sentence-transformers, scikit-learn
CV	torch, torchvision, PIL
ML	xgboost, lightgbm, scikit-learn
Infra	joblib, pandas, numpy, tqdm

Results

Model	Validation SMAPE	MAE ($)
XGBoost	18.5%	12.5
LightGBM	19.2%	13.2
Random Forest	20.1%	14.1

Stacking Ensemble	37.8%	11.9

Final SMAPE: 37.8%
Mean Absolute Error: ~$11.90
Goal Achieved: < 40% SMAPE target

Insights

- BERT embeddings outperform traditional TF-IDF in textual comprehension.
- Visual cues (color, packaging, branding) significantly influence price predictions.
- Domain-specific features like weight, brand, and premium indicators are strong predictors.
- Ensemble learning improved consistency and reduced error variance.

Conclusion

This solution delivers a multimodal ensemble model achieving approximately 37.8% SMAPE using a two-day sprint implementation. By integrating text semantics, visual perception, and domain-specific features, it sets a benchmark for intelligent product pricing systems.