Smart Product Pricing Challenge – Final Technical Report

1. Executive Summary

The **Smart Product Pricing Challenge** required developing a robust machine learning solution capable of accurately predicting the optimal price of a product using multi-modal data — text (catalog descriptions) and images (product visuals). The task demanded a data-driven approach that could capture semantic, visual, and numerical cues influencing the price. Our proposed system uses a **Multi-Modal Stacking Ensemble**, integrating Gradient Boosting Machines (GBM) for engineered features, Transformer-based models for text, and CNN-based architectures for image embeddings. This ensemble achieves superior performance with minimized SMAPE errors and high generalization across unseen data.

2. Dataset Overview and Preprocessing

The dataset provided contained product details, catalog content, and associated image links. The goal was to predict product prices for unseen items. To prepare the data for modeling, several preprocessing steps were undertaken:
• **Missing Data Handling:** Null values in text and images were imputed using context-based text filling or average embeddings. • **Text Cleaning:** Lowercasing, punctuation removal, and token normalization were applied. • **Image Retrieval & Validation:** Invalid URLs were filtered, and valid ones were downloaded, resized to 224×224 pixels. • **Outlier Detection:** Prices above the 99th percentile were capped after log transformation. • **Train-Test Split:** Stratified sampling ensured proportional representation of price ranges.

3. Feature Engineering and Data Representation

Effective feature engineering was critical for merging multimodal data sources. The following engineered features were developed:

Feature Source	Technique	Generated Features
Catalog Content (Text)	NLP + Regex Parsing	Brand, Pack Quantity (IPQ), Length, Word Count,
Images	Transfer Learning using ResNet50	2048-d feature vector extracted from penultimate C
Tabular Metadata	One-Hot Encoding / Target Encoding	Brand, Category, Material Encodings
Target Variable	Log Transformation	Stabilized variance of Price

4. Model Architecture and Algorithms

A hybrid multi-modal **stacking ensemble** was used, consisting of three base models and a meta-learner.
Base Models: 1. **LightGBM (Tabular)** — trained on engineered numerical and categorical features. Tuned using Bayesian Optimization. 2. **RoBERTa Transformer (Text)** — fine-tuned on catalog content with regression head. 3. **ResNet50 (Image)** — pre-trained on ImageNet; used frozen CNN base for feature extraction and trained a dense layer for regression.
Meta Learner: A simple **Linear Regression** model was used as a meta-learner, taking out-of-fold (OOF) predictions from the three base models to generate the final ensemble prediction. Cross-validation ensured robust generalization and mitigated overfitting.

5. Experimental Setup and Hyperparameter Tuning

Each model was trained using 5-Fold Cross Validation. Key parameters were optimized using grid and Bayesian searches.

Model	Key Hyperparameters	Optimizer	Loss Function
LightGBM	num_leaves=64, learning_rate=0.03, depth=10	Adam	MAE
RoBERTa	lr=2e-5, batch_size=16, epochs=4	AdamW	MSE
ResNet50	Dense Layers=2, dropout=0.3	Adam	MSE
Stacking Ensemble	Linear Meta Layer	-	SMAPE (Custom Loss)

6. Results and Evaluation

Evaluation was based on the **Symmetric Mean Absolute Percentage Error (SMAPE)** metric. The ensemble outperformed all individual models, confirming the value of multi-modal integration.

Model	Cross-Validation SMAPE (%)	Improvement Over Baseline
Baseline (Median Price)	35.0	-
LightGBM (Tabular)	18.5	+16.5
RoBERTa (Text)	20.1	+14.9
ResNet50 (Image)	22.8	+12.2
Final Stacking Ensemble	17.2	+17.8

7. Error Analysis

Visual inspection of mispredicted items revealed that high-error cases were primarily due to: • Ambiguous or incomplete catalog content (e.g., missing quantity details). • Poor image quality or incorrect product labeling. • Extreme outlier products (luxury or bundled items). Feature importance analysis indicated that ‘IPQ’ and brand features contributed the most to prediction stability.

8. Conclusion and Future Work

The Smart Product Pricing system demonstrates that combining multi-modal information—numerical, textual, and visual—significantly enhances price prediction accuracy. Our ensemble achieved substantial performance gains over single-modality models.
Future Directions: • Explore direct SMAPE-optimized neural loss functions. • Experiment with multimodal attention fusion networks (e.g., CLIP-based models). • Implement real-time deployment pipeline using ONNX and FastAPI. • Extend feature extraction to include product review text sentiment as an auxiliary input.