**Smart Product Pricing Challenge - Solution Overview**

**Methodology**

Our approach predicts optimal product prices by leveraging both the textual content and visual attributes of each item. We integrated advanced deep learning models suited for multimodal learning and fast, resource-efficient inference.

**Model Architecture & Algorithms**

* **Text Features**: Product title, description, and Item Pack Quantity were represented using the DistilBERT transformer, producing high-quality semantic embeddings for each product’s catalog content.
* **Image Features**: Product images were processed in real-time using a pretrained ResNet18 CNN, extracting rich image embeddings that capture packaging cues and brand characteristics.
* **Fusion & Regression**: The text and image embeddings were concatenated and passed through fully connected layers to estimate product price using regression. The model outputs a continuous positive value per sample.

**Feature Engineering**

* Cleaned and standardized catalog\_content (text) to remove artifacts and ensure consistency.
* Extracted relevant details using tokenization and handled missing/invalid images with placeholders for robust training.
* Normalized image features using standard ImageNet statistics to match pretrained CNN input expectations.

**Training Procedure**

* Employed on-the-fly image fetching from public URLs without pre-downloading, optimizing for real-world memory and storage constraints.
* Used early stopping and validation split to maximize generalization and avoid overfitting.
* Ensured predictions were strictly positive using post-processing clamps.

**Other Implementation Details**

* Custom PyTorch Dataset class for dynamic multimodal data generation.
* DataLoader configured for synchronous loading (num\_workers=0) for compatibility with Windows environments and Jupyter notebooks.
* Final predictions exported to test\_out.csv matching the sample output file format.

**Evaluation Metric**

* All experiments and leaderboard submissions followed the provided SMAPE (Symmetric Mean Absolute Percentage Error) metric for accuracy assessment.
* Validation SMAPE reported, continuous optimization for lowest possible score.

**Compliance**

* No external price lookups, web scraping, or use of outside market data.
* Models and code adhere to MIT/Apache 2.0 licensing requirements.
* Only resources and datasets provided by organizers were used