**Amazon ML Challenge 2025: Smart Product Pricing Solution**

**Team Name:** Alchemy  
**Team Members:** Lydia Dondapati,Sravani Thota,Lakshmi Lavanya Varanasi,Keerthi Sri Varsha Mudunuri  
**Submission Date:** 13 Oct 2025

**1. Executive Summary**

We developed a multimodal product pricing model integrating both textual and visual features. Product descriptions and titles were processed using advanced NLP embeddings, while product images were preprocessed and encoded via CNN features. These representations were fused to predict optimal product prices using an ensemble regression model, achieving robust SMAPE performance on validation data.

**2. Methodology Overview**

**2.1 Problem Analysis**

The dataset consisted of product details (title, description, image, and quantity).  
Exploratory analysis revealed that price correlates strongly with brand, quantity (IPQ), and product type. Outliers and non-standard price scales were detected and addressed using log transformations and scaling.

**Key Observations:**

* Item Pack Quantity (IPQ) was a key numeric indicator of price.
* Descriptive richness and specific brand mentions influenced pricing.
* Image color patterns often hinted at product category and luxury level.

**2.2 Solution Strategy**

**Approach Type:** Hybrid Multimodal Ensemble  
**Core Innovation:** Combined text embeddings from TF-IDF/BERT with CNN-based image embeddings (ResNet-50). The features were concatenated and passed through a gradient boosting regressor (XGBoost/LightGBM) for price prediction.

**3. Model Architecture**

**3.1 Architecture Overview**

**Text (catalog\_content) ─▶ NLP Embedding (TF-IDF/BERT)**

**Image (image\_link) ─▶ CNN (ResNet-50) Feature Extraction**

**↓**

**Concatenation + Feature Fusion**

**↓**

**Ensemble Regressor (LightGBM)**

**↓**

**Predicted Price**

**3.2 Model Components**

**Text Processing Pipeline:**

* Preprocessing: Cleaning text, removing stopwords, lemmatization
* Model Type: TF-IDF vectorizer + Linear Regression / BERT embeddings
* Key Parameters: max\_features=10,000, ngram\_range=(1,2)

**Image Processing Pipeline:**

* Preprocessing: Resize (224×224), normalization, RGB scaling
* Model Type: Pretrained ResNet-50 (torchvision)
* Key Parameters: Layer freezing, average pooling to 2048-dim features

**4. Model Performance**

**Validation Results:**

* SMAPE Score: **59.2720%**

**5. Conclusion**

Our hybrid model effectively captures multimodal patterns in product data, outperforming single-source baselines. The use of image and text fusion enhances price generalization for unseen categories. Future improvements include fine-tuning CNNs and transformer-based text models.

**Appendix**

**A. Code Artefacts:**  
📁 [Drive Link Placeholder – Add your Drive link here]

**B. Additional Results:**  
Feature importance plots indicate textual features contributed ~65% of the predictive signal, while image embeddings added valuable category-specific context.