

Team Name: RasMLAI MNNIT

Team Members: Anaranyo Sarkar (Leader), Aayush Kumar, Abhishek Kumar, Rushan Dayma

Submission Date: October 13, 2025

1. Executive Summary

Our hybrid LightGBM model predicts prices using a multimodal approach, combining EfficientNet image features and TF-IDF text features.

2. Methodology Overview

2.1 Problem Analysis

Initial baseline models (e.g., Linear Regression) failed to learn meaningful patterns, predicting only the average price. This confirmed a more powerful, non-linear approach was required.

2.2 Solution Strategy

We employed a hybrid model, using a pre-trained CNN for visual features and TF-IDF for text, with a gradient boosting model for the final regression.

3. Model Architecture

3.1 Architecture Overview

A two-tower pipeline extracts image features (EfficientNet) and text features (TF-IDF/IPQ) separately. The combined feature vector is then used to train the final LightGBM regressor.

3.2 Model Components

Text Processing Pipeline:

- **Preprocessing:** [Robust parsing, IPQ extraction, Log-price transform]
- **Model type:** [TfidfVectorizer]
- **Key parameters:** [max_features=15000, ngram_range=(1, 2)]

Image Processing Pipeline:

- **Model type:** [EfficientNet-B0 (from timm)]
- **Key parameters:** [Pre-trained, used as feature extractor]

4. Model Performance

4.1 Validation Results

- **SMAPE Score:** 52.9658% (on a 20% hold-out).

5. Conclusion

The final model is a robust, multimodal pipeline that successfully fuses visual and text data. Future work will focus on using Transformer embeddings and hyperparameter tuning to further reduce the SMAPE score.