Product Length Prediction

Bhanu Prasanna Koppolu

Table of contents

What I Built

Built a machine learning system to predict Amazon product physical length from text metadata using a multi-embedding ensemble approach. Combined 4 different text embeddings (MiniLM, MPNet, DistilUSE, E5-Small) with KNN retrieval features and product type embeddings in a neural network.

What I Learned

Loss function choice matters enormously. Switching from Huber loss (94% MAPE) $\rightarrow$ Direct MAPE optimization (59%) $\rightarrow$ MAPE with log-target transform (51.78%) reduced error by 42 percentage points. The key insight: train for what you're measured on.

Also learned that pre-computing embeddings and using simple MLP architectures often outperform complex models when the data preprocessing and loss function are properly optimized.

Project

[GitHub Repository](#)

Final Performance: 51.78% MAPE (48.22 competition score)