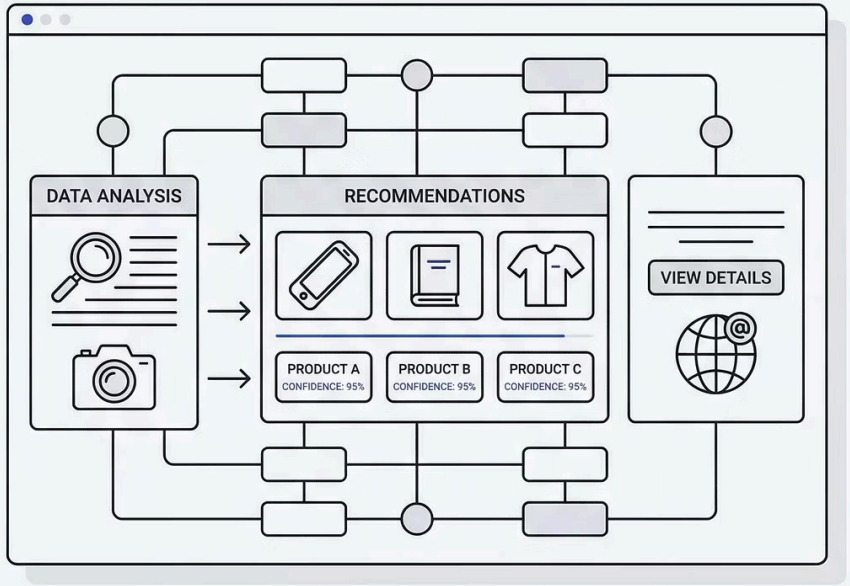
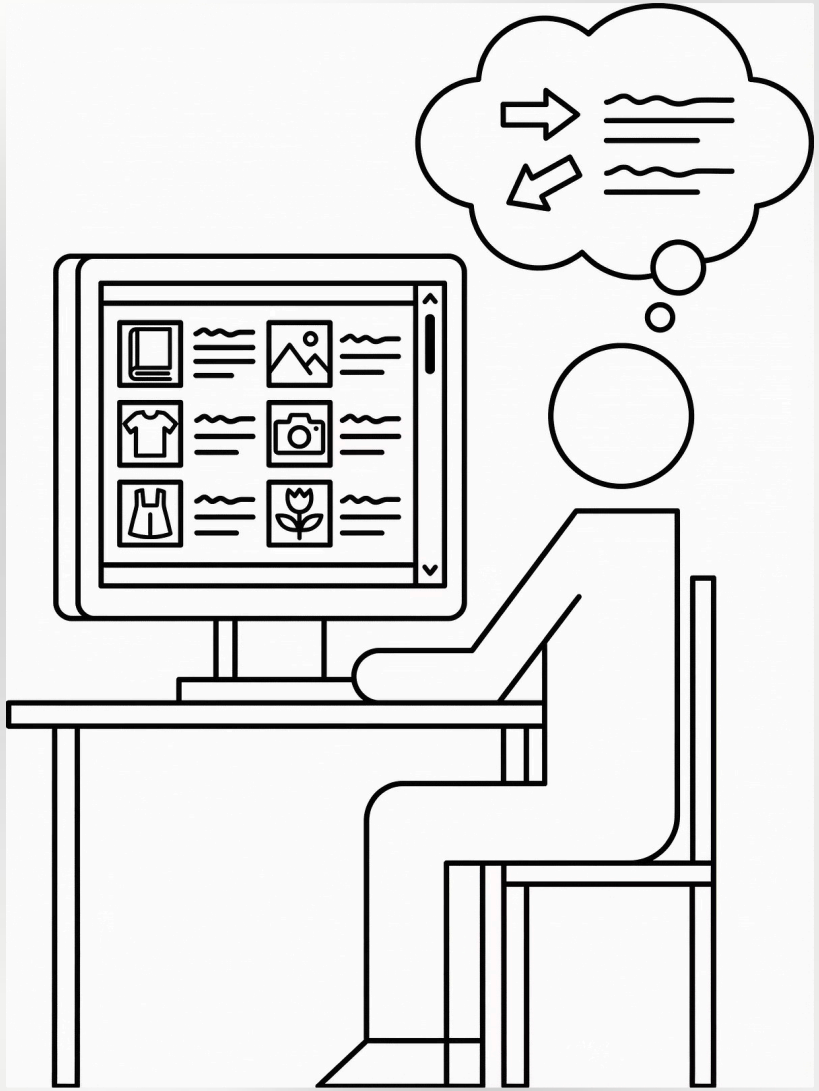


Product Recommendation with Multimodal Siamese Networks



The Challenge: Beyond Simple Recommendations

The Problem with Unimodal Systems

Traditional recommendation systems often rely on a single data type, like text descriptions or image features. This leads to an incomplete understanding of product similarity and less accurate recommendations for users.

The Need for Richer Data

Customers consider both visual and textual aspects when making purchasing decisions. Our goal is to leverage this human intuition to create a more sophisticated and effective recommendation engine.

Our Solution: Multimodal Siamese Networks

We developed an innovative approach that combines the power of natural language processing and computer vision to understand product similarity more comprehensively.

Synergistic Data Analysis

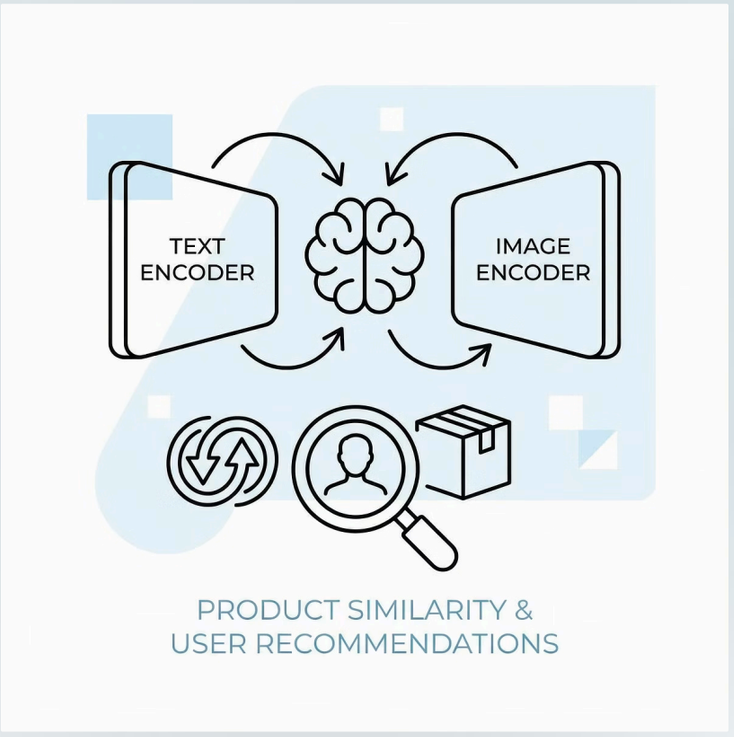
By integrating both text and image data, our model captures a richer context, leading to highly relevant product suggestions.

Enhanced Similarity Metrics

The Siamese network architecture learns robust similarity embeddings, allowing for nuanced comparisons between products even with subtle differences.

Improved User Experience

Ultimately, this leads to more accurate and satisfying recommendations, boosting user engagement and potentially increasing sales.



Dataset & Foundation: Amazon Product Data

Leveraging a Rich Dataset

Our project utilizes the vast and diverse **Amazon Product Dataset**, providing a robust foundation of product descriptions and corresponding images.

Extracting Key Features

This dataset allows us to train our multimodal model on a wide array of product categories, ensuring broad applicability and accuracy.

Real-World Relevance

Working with real-world e-commerce data ensures our solution is practical and scalable for commercial applications.

Model Architecture: The Multimodal Siamese Network

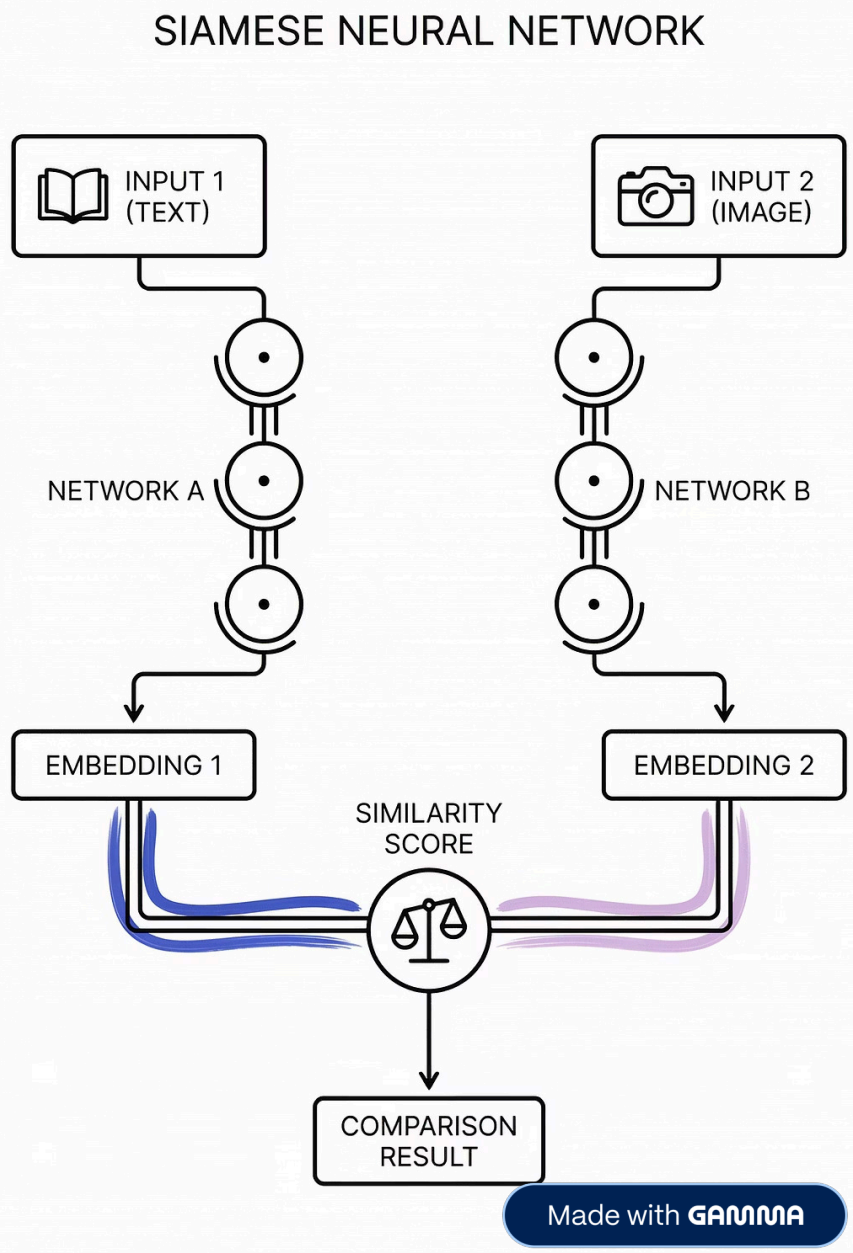
At the core of our solution is a sophisticated Siamese neural network designed to process and compare different data types.

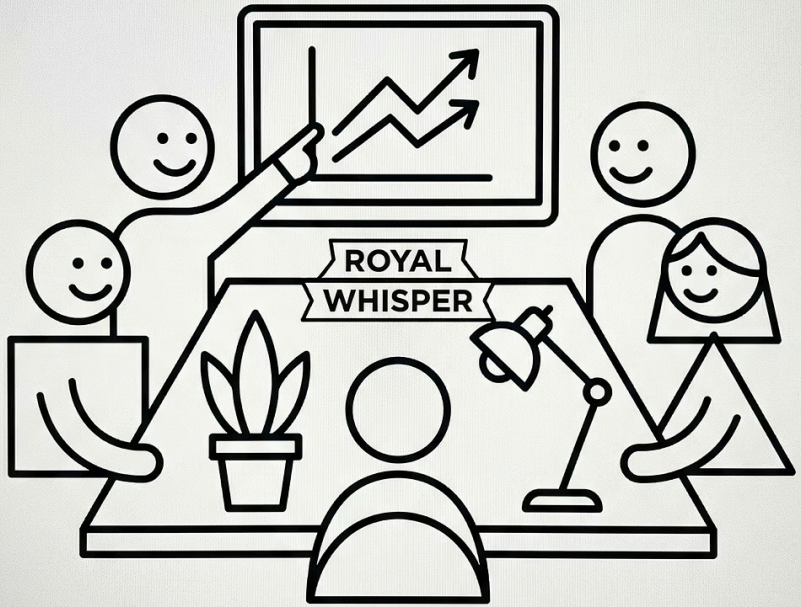
Text Encoder: LSTM

- Processes product descriptions and extracts semantic features.
- Captures contextual information and nuances in language.

Image Encoder: CNN

- Analyzes product images to identify visual characteristics.
- Learns to differentiate between similar and dissimilar visual patterns.





Thank You!

Questions & Discussion

We appreciate your time and welcome any questions you may have about our project.