



SKU Matching Across Fragmented Retail Data

Services: **NLP and text analytics**



Context

A retail analytics workflow required aligning product-level transaction data from informal vendors with a central product catalogue. Source data was highly unstructured, with inconsistent naming conventions, missing identifiers, and significant variation in how products were recorded across vendors. This made reliable aggregation and analysis difficult, particularly when integrating data across multiple sources.

Solution

Developed an NLP-driven matching system that transforms raw product descriptions into structured attributes (e.g., brand, variant, pack size, weight) using a fine-tuned language model. These structured representations were then used within a scoring-based entity resolution framework to match products against a master catalogue, with confidence scoring and fallback logic to handle incomplete or ambiguous cases. This shifted the problem from fragile string matching to a more robust, attribute-based approach.

Impact

Improved matching accuracy and consistency across fragmented data sources, while eliminating reliance on brittle rule-based approaches. The solution enabled scalable integration of new vendor data and more reliable downstream aggregation and reporting. It also reduced manual intervention required to maintain matching logic as new product variations emerged.