

UPCS Automated Classification Rules

This file describes a simple, explainable rules-first approach for auto-suggesting UPCS categories from SKU text.

Inputs

- sku_name
- sku_description (optional)
- brand (optional)
- ingredients (optional)

Process (recommended)

1. Normalise text: lower-case, strip punctuation, normalise whitespace.
2. Apply rules in priority order:
 - Exact phrase rules (e.g. "fresh pasta")
 - Regex keyword rules (e.g. "\bcrisps\b")
3. Produce top-N suggestions with confidence scores.
4. Human review required below a confidence threshold (e.g. < 0.75).
5. Log outcomes to improve rules and later train an ML model.

Outputs

- suggested_uos_category_id
- confidence
- matched_rule_id
- matched_text_span (optional)

See: `upcs_nlp_classification_rules.csv`.