Product Catalog Matching System

TAKE-HOME ASSIGNMENT

CHARALAMPOS TSIOLIS

26 NOVEMBER 2025

Executive summary

Developed an automatic system to match unstructured product descriptions to the correct SKU

On validation data , reached

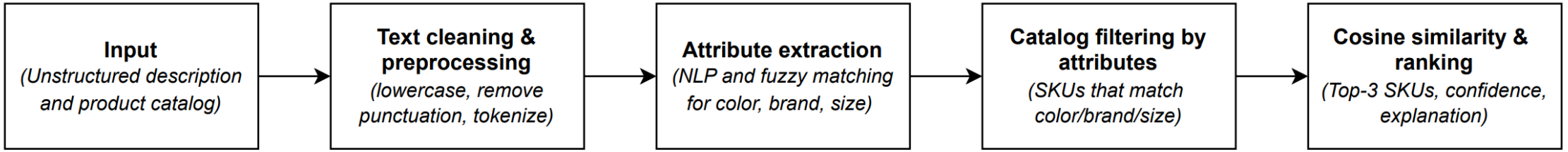
87% Top-1 & 93% Top-3 accuracy

Attributes (color, brand, size) are used to improve matching consistency

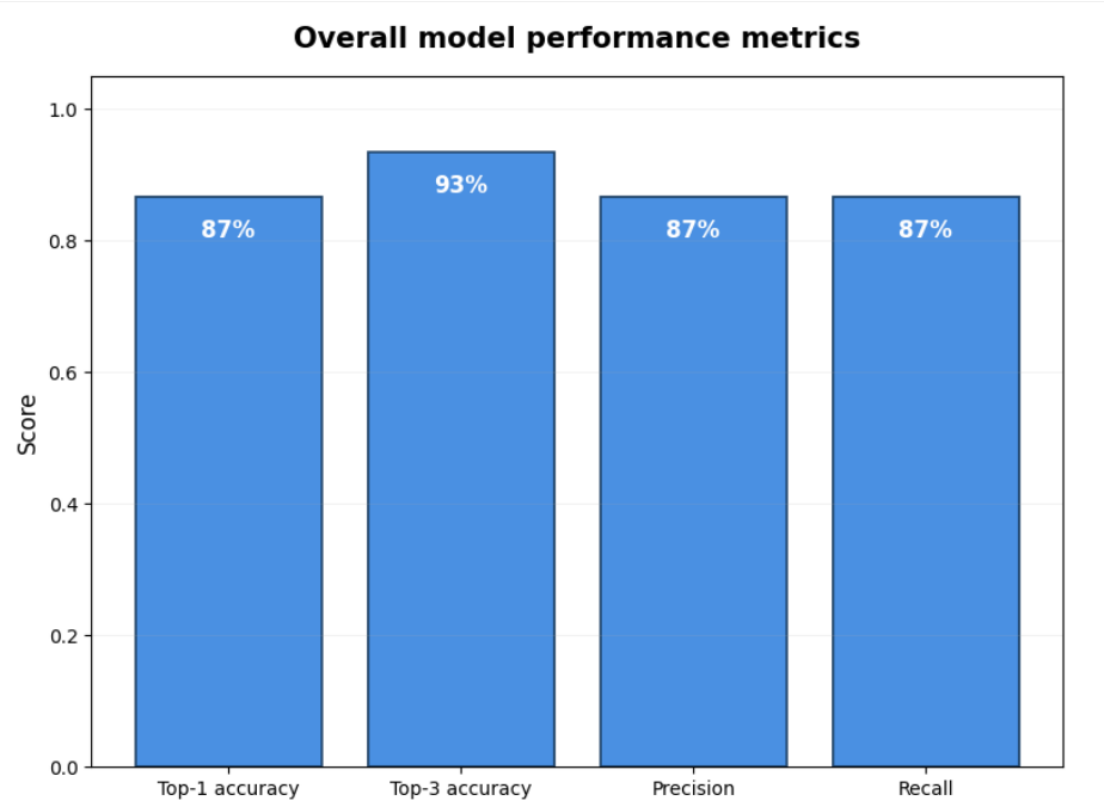
Saves/minimizes time spent by approximately 2-3 hours/day

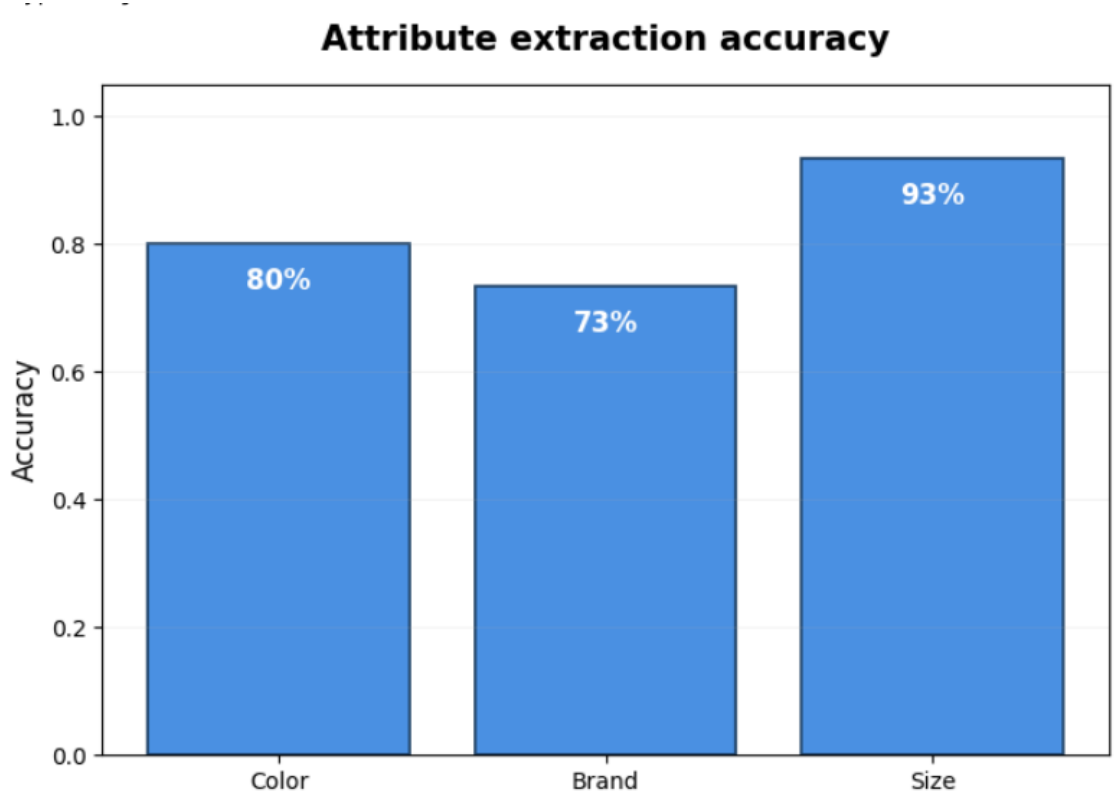
Model overview

- Preprocessing of the text and TF-IDF vectorization
- NLP with fuzzy matching (color/brand/size)
- Filtering of catalogs based on these attributes
- Ranking SKUs with cosine -> gives Top-3 matches , confidence and explanation



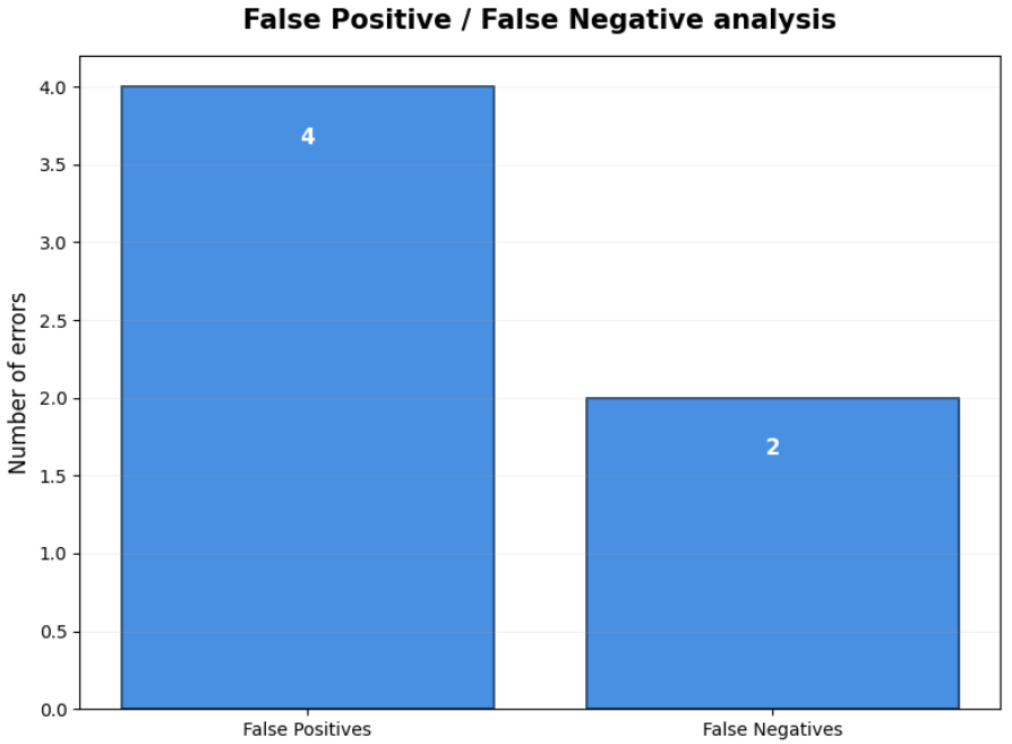
Model performance





Error analysis

- Incorrect matches scored much less in terms of confidence
- Some mismatches arise when there is no brand in the descriptions. Or when common words (like “style”) are mistaken for attributes



Business recommendations & ROI

01

Auto-approve → with confidence greater than 0.75

02

Automation of matching about 90% a day saves about 2 to 3 hours of work a day

03

Extracting more attributes from the text (e.g. category, season , material) to improve matching accuracy