**Product matching**

Presented by Ahmed Sief eleslam ahmed

15/2/2025

**Problem Statement:**

Isupply platform faces challenges in matching products from different sellers due to variations in naming and pricing.

**Objective:**

Develop an automated method to match products using a combination of text similarity and price comparison.

Enhance User Experience by quickly matching

Faster Data review with high accuracy

**Data Collection & Preprocessing**

* Master Product List (Official Product Names & Prices)
* Seller Dataset (Seller’s Product Names & Prices)

**Data Cleaning Techniques:**

* Standardizing Arabic text
* Removing unwanted words (e.g., discounts, promotions)
* Handling missing values & formatting numbers

**Matching Methodology**

As we found that the price is more unique due to dosage interference in text so we decided to make a grid search to get the perfect distribution of weights so,

* Text Similarity (25%)
  + TF-IDF vectorization (Character n-grams: 2-4)
  + Cosine Similarity for text matching
* Price Similarity (75%)
  + Formula: 1 - abs(price1 - price2) / max(price1, price2)
  + Normalization to ensure fair contribution
* Final Score Calculation:
  + Combined Score = (0.25\* Text Similarity) + (0.75\* Price

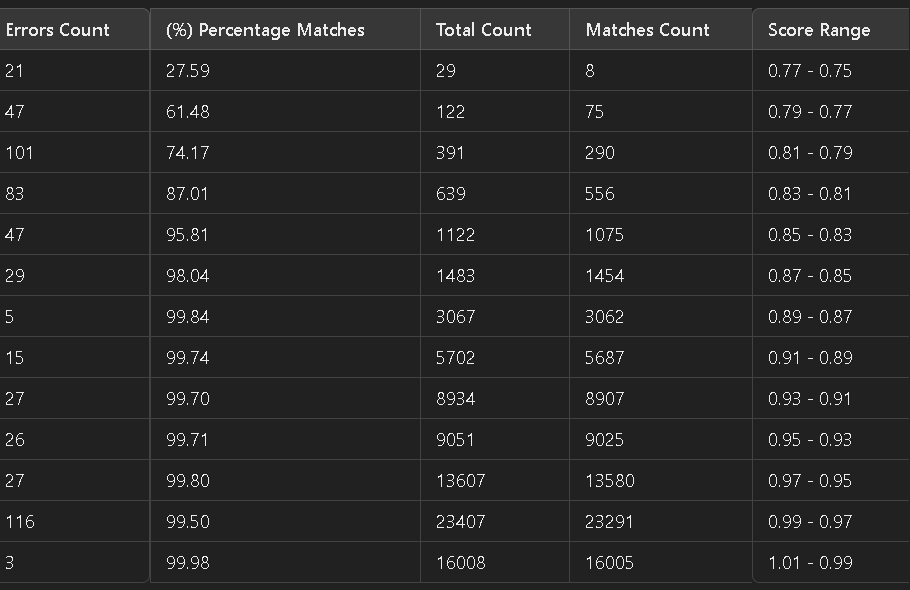
Similarity)

**Implementation & Workflow**

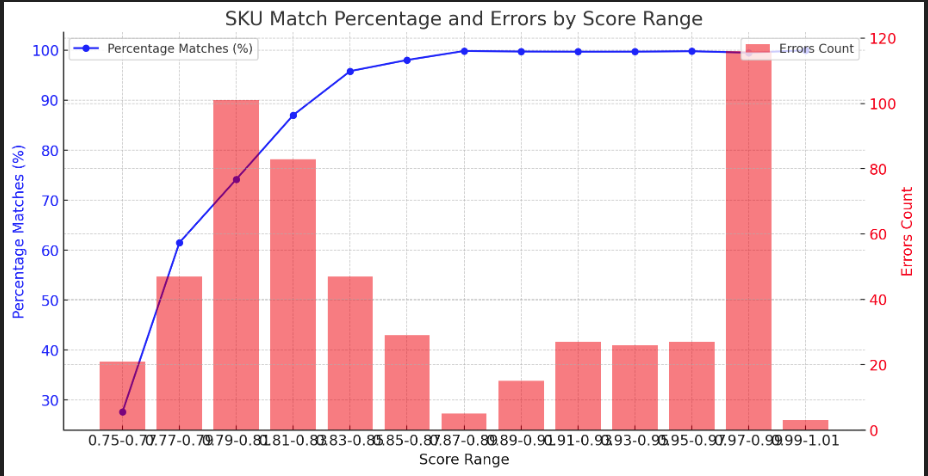
* Pipeline Overview:
  + Load & preprocess data
  + Compute text similarity (TF-IDF + Cosine Similarity)
  + Compute price similarity
  + Generate final matching scores
  + Filter & evaluate matches
  + Extract a matched data file

**Performance Analysis**

The model has a 99.3% a total accuracy labeled by the true sku

The distribution of correct matches according to sku and scores is:note: this error counts at the worst cases some skus where wrong in the dataset

And there is graph representation



**Challenges & Solutions**

* **Challenges:**
  + Variations in product naming conventions
  + Handling promotions & misleading text
  + Different currency formats & price fluctuations
  + Algorithms mixing between embedded prices and dosages
  + Make a good confidence score
* **Solutions Implemented:**
  + Enhanced text preprocessing
  + Adjusted weights for similarity calculation
  + Used SKU-based ground truth for validation

**Conclusion & Future Work**

* Conclusion:
  + Effective method for automated product matching
  + Balanced approach using text and price similarity
  + Need for further fine-tuning on datasets
* Future Work:
  + Incorporating deep learning (BERT for Arabic text)
  + Improving handling of synonyms & abbreviations
  + Expanding dataset for better generalization

**Thank you!**

* + LinkedIn: [https://www.linkedin.com/in/ahmed-sief-eleslam-124b4a249/]
  + Email: [ahmedsiefeleslam@gmaill.com]
  + GitHub: [https://github.com/ahmedsief0/Product-Matching]