competera | Whitepaper

Manual vs. Automated Matches

Product Comparison as an Essential Retail Element

Whether a retailer monitors competitors to make sure its prices are optimal or uses an aggressive pricing strategy to win the market share, accurate data is very important.

It's hard to estimate the number of data variables retailers need to consider. One of them—mostly underestimated—are the product matches. Percentage of comparisons is one of the primary indicators of data quality; indirectly it affects customer loyalty, market share, and therefore, margins.

There are several components successful, or unsuccessful, matches consist of: the name of the product (title), product description with attributes (fixed or unfixed set of values), its picture and price. Each of these components may vary for each product by different retailers.

Let's say the retailer has to set a price for Pixel 2 XL smartphone. With incorrect product matches, he might set its prices based on competitors' refurbished products, different storage settings, color, etc. Thus, the prices would not be optimal from a customers' perception. Therefore sales and overall retailer's metrics would suffer.

The Real Cost of Tiny Mistakes

At the beginning of the pricing journey retailers often start to monitor and match a limited part of their assortment, around 1K, mostly KVI-products. Then, when the pricing process is set, a retailer usually begins to monitor a more significant portion of the range.

It is an area where matches start to play a huge role in pricing. Even a small mistake leads to enormous losses. For enterprise retailers which sell thousands of products, it's even more critical to maintain the margin on every product, with no exception. Therefore, accurate matches are crucial.

To describe the statement mentioned above, let's count the number of pricing mistakes a retailer can make with incorrect matches. If a retailer sells 50K products, and the accuracy of matches is 70%, there is a possibility that 15K of products are poorly priced.

This is a reason retailers—especially enterprises—are looking for solutions that provide 98% of accurate comparisons or more.

There are only two approaches to match products: manual (visual) and automated matches. Each one of them has its advantages and drawbacks.

Accuracy of matches	Products affecting retailer's margin
70%	15K
80%	10K
90%	5K
95%	2.5K

Automated Matches

The main advantages of the automated matching process are its price and speed.

Of course, it depends on the technology used for matching and number of product identifiers, but the speed is the main advantage of automated matches. Machine learning algorithms can match 1K products per minute which is far more than the capabilities of the manual matching process. Therefore, the speed of matching leads to the price decrease.

Also, every ML algorithm shows a presumption of matching correctness. It shows that the probability is, let's say, 63% or 90% for a particular match. This presumption shows if there is a need to re-check data or not.

In the meantime, automated matches have all the disadvantages every machine learning algorithm has. It depends on data quality and size of data set:

Automated Matches

If the algorithm was trained on particular data set, and data shift happens (e.g., competitor changes the name of products), data shift point needs to be detected and the whole scope of products has to be rematched with human involvement.

To complete the picture, you need to mention an issue automated matches share with the manual ones. The quality of matches vary by industry and product category. For the laptop category, where predefined product identifiers exist, the matches accuracy metric will be higher than 90%. In the meantime, for groceries, the accuracy will be lower as long as retailers tend to be more liberal with product names, error percentage increase and the same metric will be around 80%.

Automated matches are best suited for sites with fixed identifiers (UPC, GTIN, ISBN), e.g. Google Shopping. In the event a retailer has a unique identifier aligned with such website, automatches can provide high accuracy. In the meantime, this approach suffer from less regulated identifiers or direct retailer websites, where only manual matches can be used.

Manual Matches

Unlike the automated ones, manual matches provide quick launch, ease of scale, and make the capacity planning possible. When built correctly, they give a retailer more accurate data, though it's more expensive.

Manual matches use different algorithms to provide the retailer with reliable data. It could be an hierarchical structure where the matching team matches required products, then team leads check results, after that, another person responsible for data quality audits the job of the team leads. Or it could be an ABC-flow when two person (A and B) match the same product: if their results are
similar, the product is matched; if not, the same product goes to the third person (C).

Algorithms of manual matching described above are just tip of the iceberg, there are much more options to deliver far more than 90% of qualified data in the scope.

The main disadvantage of the manual matches—besides the price—is their speed. The average capacity of a matcher is around 600 matches per day. Also, there is always a chance of human mistake.

Manual Matches

It's important to wisely choose the process of matching, or the data provider.

Some providers ask random people to do matches, e.g., through online services "earn-and-go." It's a time consuming and full of error practice, which leads to low data quality.

The more efficient approach is to hire and teach the professional matching team. This option is more accurate and provides a higher matching speed. A data provider with such a team is agile enough and can deliver a huge number of matches per day.

Another advantage of manual matches, compared to automated mentioned above is the capability to work with "complex" industries, those where products have many optional identifiers. Perfumes, toys, apparel... and the list goes on.

Manual matches are the best option when a retailer needs to get correct product data as soon as possible. Additionally, visual comparison of products is possible even for extremely secured or complex competitor websites.

Disclaimer

Manual matches rarely solve the issue of private or white label products.

Yet, private labeled products have more significant margin than A-branded products, and their pricing can be margin-driven instead of competitive-based.

Therefore, a retailers' priority in matching products with competitors tilts toward A-branded products where it is more commercially viable and aligns with A-branded products market pricing strategy.

In the meantime, if there is a need for private label matches, e-commerce businesses can choose more profitable and top seller products (100-200 products per category manager) and match them frequently with little effort on their own.

Capacity of Manual and Automated Matches

	Automated matches	Manual matches
Avg % of correct matches*	30-70%	90-98%
Speed	1K matches/min	600 matches/day by 1 person
Time to set	3-4 weeks to train the algorithm	up to 24 hours for a dedicated matching team
Data requirements	• Healthy competitive data • Large data set • Machine learning engineers	• QA procedures • Constant team training

The most problematic industries for both approaches are building materials, fashion, jewelry, toys, and other product categories with identifiers described by a picture and unstructured text.

This issue comes from approach each store use to describe the specifications of the product, e.g., price per square meter or 1 L of paint. Because of the human involvement, manual matches handle these categories far better than automated.

Matches Accuracy Benchmark by Industry

	Automated matches	Manual matches
Apparel/Footwear	40-70%	85-95%
Consumer Electronics	90-97%	90-98%
Tools	65-80%	90-98%
Automotive supplies	70-85%	85-92%
Books	95%+	80-93%
Home improvement	70-85%	80-90%
Health & Beauty	80-90%	85-92%
Sporting Goods	75-90%	85-92%
Toys & Hobbies	50-85%	85-95%
Building	40-70%	75-90%
Office Supplies	70-90%	87-95%
Grocery	30-50%	80-90%
Jewelry	30-60%	75-95%
Pet Supplies	70-85%	88-96%
Baby Products	80-90%	90-95%

The Best Option is to Merge These Two Approaches

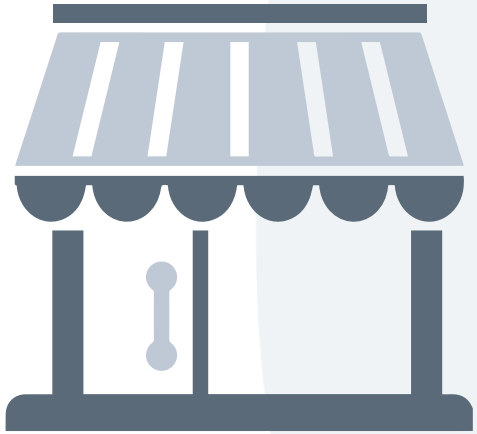
To get correct matches at the speed of the changes, a retailer needs to combine the best of two methods: set the process of automated matching and enhance it with the accuracy of manual matches.

For example, a manual check to qualify automated matches can (a) improve the matching algorithm, and (b) provide precise quality of matches. With such a mixed approach, a high percentage of comparisons by many parameters can be delivered, along with a low amount of zero (not-found) prices and minor errors in the collected data.

Use Case on Merging Automated and Manual Matching Approaches

If the retailer in a particular region needs to monitor X number of competitors and check prices on Amazon and Google Shopping, he has an opportunity to merge automated and manual matches.

In this case, products from marketplaces are matched automatically by ASIN and GTIN identifiers. Items collected directly on competitors websites are matched manually.



The Main Question Retailers Ask

The leading question retailers are facing, is whether it's better to hire a third-party for matches, create its in-house alternative or use freelancers to do the job.

In the end, it should be a data-driven decision. A simple calculation shows that third-party solutions are the best option.

Here is a simple example of how much retailer pays for 1000 products matches from 5 competitors' websites:

	In-house	Freelancers	Third-party service
Product matches and rematches	~$1600/mo	~$400/mo	~$100/mo
Time	3 days	1 week	3 hours
Operational costs (office, appliances, supplies)	~$400/mo	No	No
Q&A procedures	Included in salary	Excluded	Included in the monthly fee
Data and matches quality	Questionable, needs additional resources	Questionable, needs additional resources	SLA guaranteed
Continuously improved matching process on different sources and data sets	No	No	Yes

The Main Question Retailers Ask

There is a chance to create an in-house solution cheaper by hiring freelancers. However, it won't help to keep data relevance for not-found or new products. Also, it's almost impossible to get any SLA commit from in-house team or freelancers. Average data quality in such scenario is below 80% (this number is based on Competera internal survey among current customers).

Even though a third-party solution is the best way to set matching flow, there are several key question retailer need to ask data providers before choosing the one to work with:

- Is there an SLA for quality of product matches or not?
- What amount of time the data provider needs to add new products (for example, 1000 items), is there an SLA timeline for this task?
- How many people are going to be involved in the matching process?
- How does the recheck process look like?
- How does a provider control the Q&A process?
- Does a provider have any clients in your industry?
- Is there any flow for incorrect matches reporting and retrieval, and whether it is included in an SLA or not?

If a retailer gets the appropriate answers to all these questions from the data provider, it means that the provider is able to deliver accurate and timely data.

Test Competera Matches Quality

Request a pilot to test the quality of data collection and product matching which Competera provides for retail winners worldwide

Request Pilot