AI-powered Product Matching: The Key to Competitive Pricing Intelligence in eCommerce

30th Aug, 2019

BY SANKET

With thousands of products and hundreds of online retailers to choose from, the average modern-day shopper usually compares prices across several e-commerce sites effortlessly before often settling for the lowest priced option. As a result, retailers today are forced to execute millions of price changes per day in a never-ending race to be the lowest priced – without losing out on any potential margin.

Identifying, classifying, and matching products is the first step to comparing prices across websites. However, there is no standardization in the way products are represented across e-commerce websites, causing this process to be fairly complex.

Here's an example:

Crazy Dog T-Shirts
Keep Calm And Drink Wine T Shirt Funny Drinking Shirt Wine Lovers Tee
★★★★★ ▼ 7 customer reviews

Women's Keep Calm and Drink Wine Cotton T-shirt
★★★★☆ ▼ 1 Reviews
Sale Starts at INR974.69

What's needed is a pricing intelligence solution that first matches products across several websites swiftly and accurately, and then enables automated tracking of competitor pricing data on an ongoing basis.

Pricing intelligence solutions already exist. What's wrong with using them?

There are several challenges with the incumbent solutions in the market – the biggest one being that they don't work in a timely manner. In essence, it's like deferring the process of finding actionable information that helps retailers acquire a competitive advantage, and instead doing it in hindsight. Like an autopsy of sorts.

Here are the various solution types we have in the market today:

→ **Internally developed systems –** Solutions developed by retailers themselves often rely on heavy manual data aggregation and have poor product matching capabilities. Since these solutions have been developed by professionals not attuned to building data crunching machines, they pose significant operational challenges in the form of maintenance, updates, etc.

→ **Web scraping solutions –** These solutions have no data normalization or product matching capabilities, and lack the power to deliver relevant actionable insights. What's more, it's a struggle to scale them up to accommodate massive volumes of data during peak times such as promotional campaigns.

→ **DIY solutions –** These solutions require manual research and entry of data. It goes without saying that due to the level of human intervention and effort required, they're expensive, difficult to scale, slow, and of questionable accuracy.

As common as it is nowadays, AI has the answer

DataWeave's competitive pricing intelligence solution is designed to help retailers achieve precisely the competitive advantage they need by providing them with accurate, timely, and actionable pricing insights enabled by matching products at scale. We provide retailers with access to detailed pricing information on millions of products across competitors, as frequently as they need it.

Our technology stack broadly consists of the following.
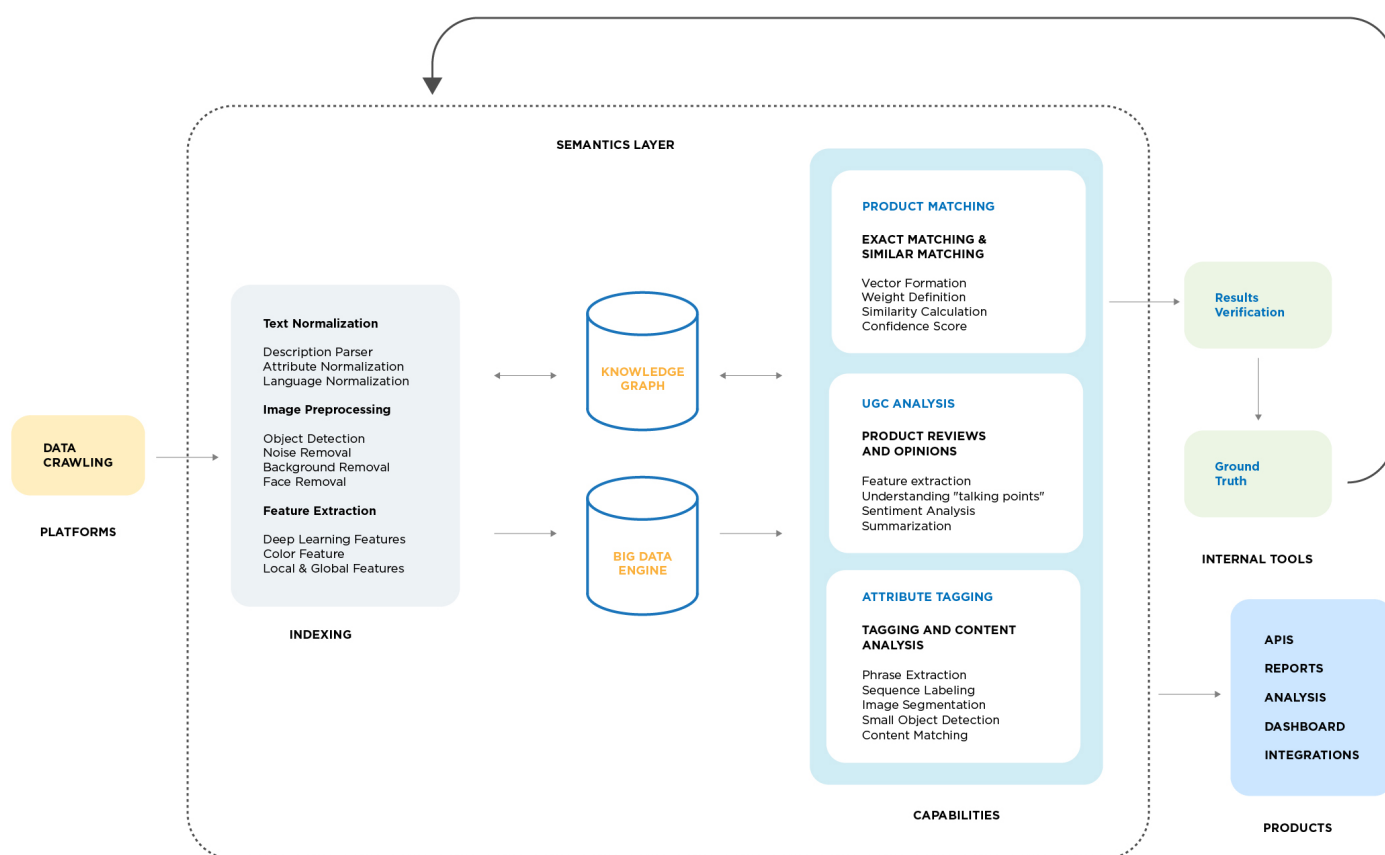
1. Data Aggregation

At DataWeave, we can aggregate data from diverse web sources across complex web environments – consistently and at a very high accuracy. Having been in the industry for close to a decade, we're sitting on a lot of data that we can use to train our product matching platform.

Our datasets include data points from tens of millions of products and have been collected from numerous geographies and verticals in retail. The datasets contain hierarchically arranged information based on retail taxonomy. At the root level, there's information such as category and subcategory, and at the top level, we have product details such as title, description, and other <attribute, value> relationships. Our machine learning architectures and semi-automated training data building systems, augmented by the skills of a strong QA team, help us annotate the necessary information and create labeled datasets using proprietary tools.

2. AI for Product Matching

Product matching at DataWeave is done via a unified platform that uses both text and image recognition capabilities to accurately identify similar SKUs across thousands of e-commerce stores and millions of products. We use an ensemble deep learning architectures tailored to NLP and Computer Vision problems specific to us and heuristics pertinent to the Retail domain. Products are also classified based on their features, and a normalization layer is designed based on various text/image-based attributes.
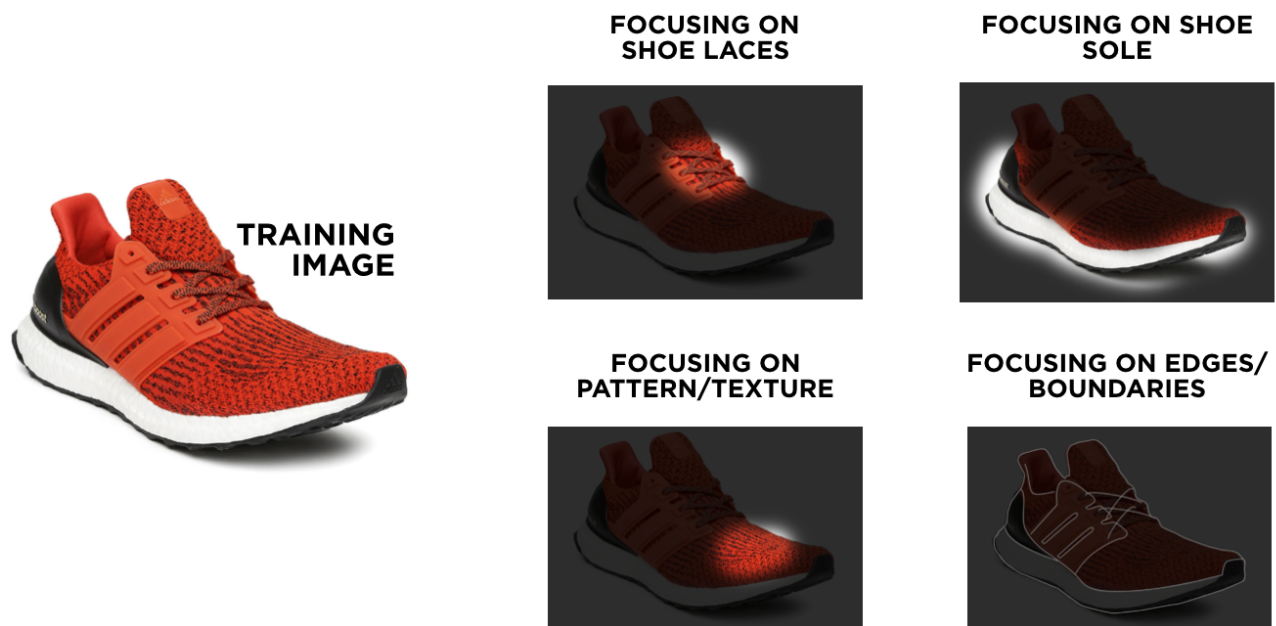
FEEDBACK LOOP

SEMANTICS LAYER

DATA CRAWLING

PLATFORMS

Text Normalization

Description Parser
Attribute Normalization
Language Normalization

Image Preprocessing

Object Detection
Noise Removal
Background Removal
Face Removal

Feature Extraction

Deep Learning Features
Color Feature
Local & Global Features

INDEXING

KNOWLEDGE GRAPH

BIG DATA ENGINE

PRODUCT MATCHING

EXACT MATCHING & SIMILAR MATCHING

Vector Formation
Weight Definition
Similarity Calculation
Confidence Score

UGC ANALYSIS

PRODUCT REVIEWS AND OPINIONS

Feature extraction
Understanding "talking points"
Sentiment Analysis
Summarization

ATTRIBUTE TAGGING

TAGGING AND CONTENT ANALYSIS

Phrase Extraction
Sequence Labeling
Image Segmentation
Small Object Detection
Content Matching

CAPABILITIES

Results Verification

Ground Truth

INTERNAL TOOLS

APIS

REPORTS

ANALYSIS

DASHBOARD

INTEGRATIONS

PRODUCTS

Our semantics layer, while technically an integral part of the product matching process, deserves particular mention due to its powerful capabilities.

The text data processing consists of internal, deep pre-trained word embeddings. We use state-of-the-art, customized word representation techniques such as ELMO, BERT, and Transformer to capture deeply contextualized text with improved accuracy. A self-attention/intra-attention mechanism learns the correlation between the word in question and a previous part of the description.

Image data processing starts with object detection to identify the region of interest of a given product (for example, the upper body of a fashion model displaying a shirt). We then leverage deep learning architectures such as VggNet, Inception-V3, and ResNet, which we have trained using millions of labeled images. Next, we apply multiple pre-processing techniques such as variable background removal, face removal, skin removal, and image quality enhancing and extract image signatures via deep learning and machine learning-based algorithms to uniquely identify products across billions of indexed products.

FOCUSING ON SHOE LACES

FOCUSING ON SHOE SOLE

FOCUSING ON PATTERN/TEXTURE

FOCUSING ON EDGES/ BOUNDARIES

TRAINING IMAGE

Finally, we efficiently distribute billions of images across multiple stores for fast access, and to facilitate searches at a massive scale (in a matter of milliseconds, without the slightest compromise on accuracy) using our image matching engine.

3. Human Intelligence in the Loop

In scenarios where the confidence scores of the machine-driven matches are low, we have a team of Quality Assurance (QA) specialists who verify the output.

This team does three things:

→ Find out why the confidence score is low

→ Confirm the right product matches

→ Figure out a way to encode this knowledge into a rule and feed it back to the algorithm

In this way, we've built a self-improving feedback loop which, by its very nature, performs better over time. This system has accumulated knowledge over the 8 years of our operations, which is going to be hard for anyone to replicate. Essentially, this process enables us to match products at massive scale quickly and at very high levels of accuracy (usually over 95%).

4. Actionable Insights Via Data Visualization

Once the matching process is completed, the prices are aggregated at any frequency, enabling retailers to optimize their prices on an ongoing basis.

Pricing insights are typically consumed via our SaaS-based web-portal, which consists of dashboards, reports, and visualizations.
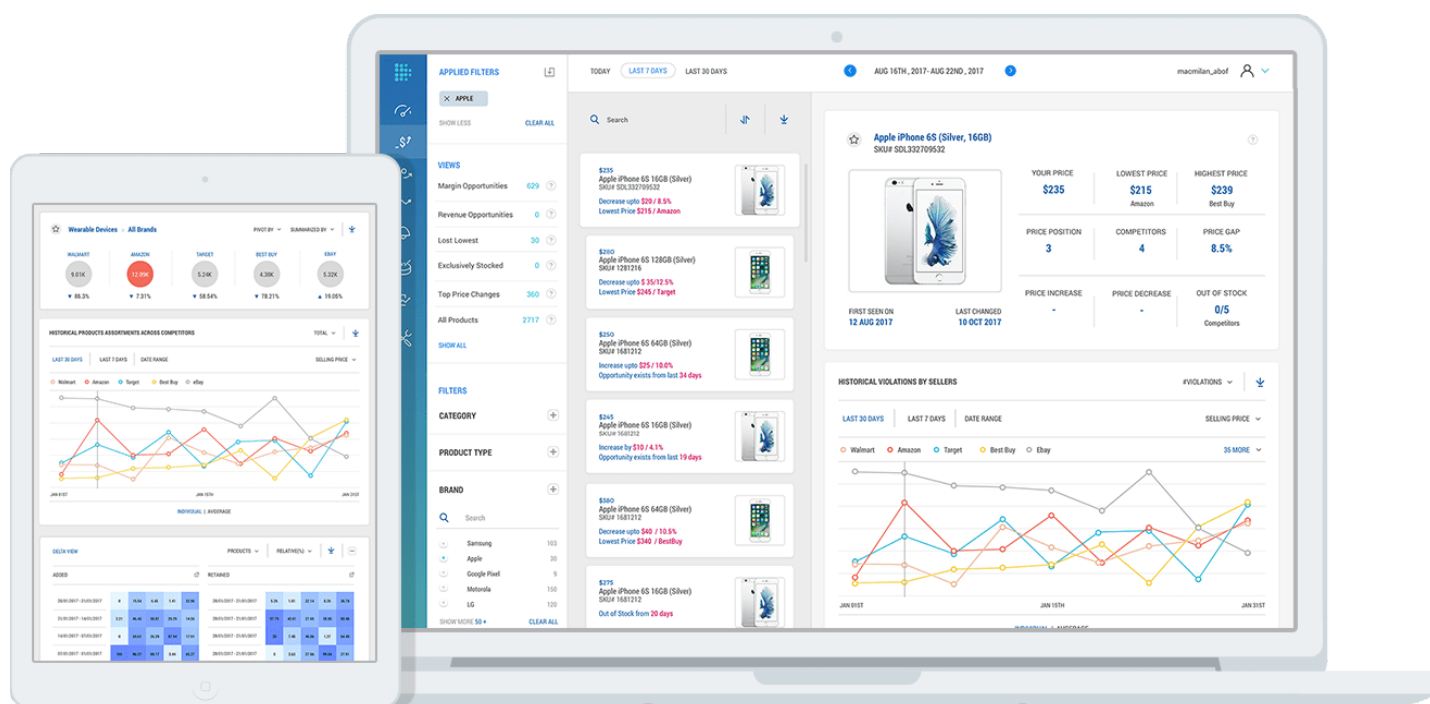
Alternatively, we can integrate with internal analytics platforms through APIs or generate and deliver spreadsheet reports on a regular basis, depending on the preferences of our customers.

To summarize

The benefits of our solution are many. Detailed price improvement opportunity-related insights generated in a timely manner empower retailers to significantly enhance their competitive positioning across categories, product types, and brands, as well as ability to influence their price perception among consumers. These insights, when leveraged at a higher granularity over the long term, can help maximize revenue through price optimization at a large scale.

Our solution also helps drive process-based as well as operational optimizations for retailers. Such modifications help them better align themselves to effectively adopt a data-driven approach to pricing, in turn helping them achieve much smarter retail operations across the board.

All of this wouldn't be possible if the product matching process, inherent to this system, was unreliable, expensive, or time-consuming.

If you would like to learn more about DataWeave's proprietary product matching platform and the benefits it offers to eCommerce businesses and brands, **talk to us now**!

- Sanket Patil

VP of Products & Data Science at DataWeave, 30th Aug, 2019

ARTIFICIAL INTELLIGENCE COMPANY E COMMERCE STRATEGY