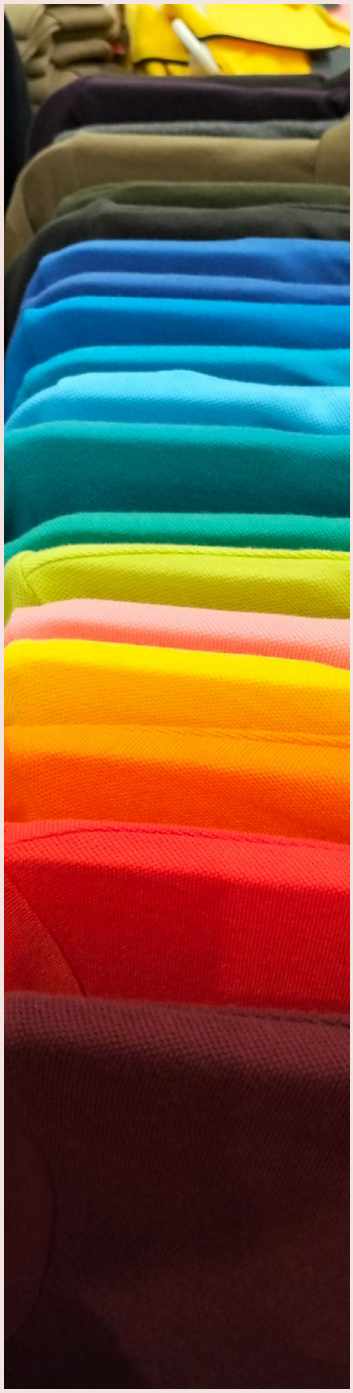
CASE STUDY



USING HUMAN INTELLIGENCE TO TRAIN THE MACHINE

Abstract

Ashley Rogers, Lead ML Engineer at a $22.3 billion American omni-channel retail giant, was setting up a model for comparing product listings across competitor sites covering 20K+ URLs and 100 product types. Recognising that it would be challenging to maintain accuracy with such a high volume ask, she looped in Infosys BPM for support. This case details how a team from Infosys BPM swiftly compared all product listings, annotated the variations, and labelled over 21,000 datasets with 100% accuracy, resulting in valuable competitor insights for the retailer and improved model performance.

Infosys®
Navigate your next

When managing machines becomes a challenge

Ashley Rogers is the Lead ML engineer at a $22.3 billion American omni-channel retail giant operating across 722 offline locations. In her role, she is responsible for overseeing the training of the company's machine learning model and managing related programs. Recently, she was tasked to monitor and improve the ML model's ability to compare product listings across competitor sites and adapt to any variations.

As a part of this project, Ashley's team were analysing and identifying product-level differences across platforms and labelling each dataset for further processing. Since the data was being used to train and evaluate the ML model, the team would have to maintain 100% accuracy throughout, ensuring that every entry was correct, verified, and consistent across multiple product categories and formats.

However, with more than 20,000+ URL's and 100 product types in scope, this was no easy ask. During her weekly check-in with the team, Ashley was quick to realise that the current data labelling model would not be sustainable in the long-run and would potentially lead to errors and inaccuracies overtime.

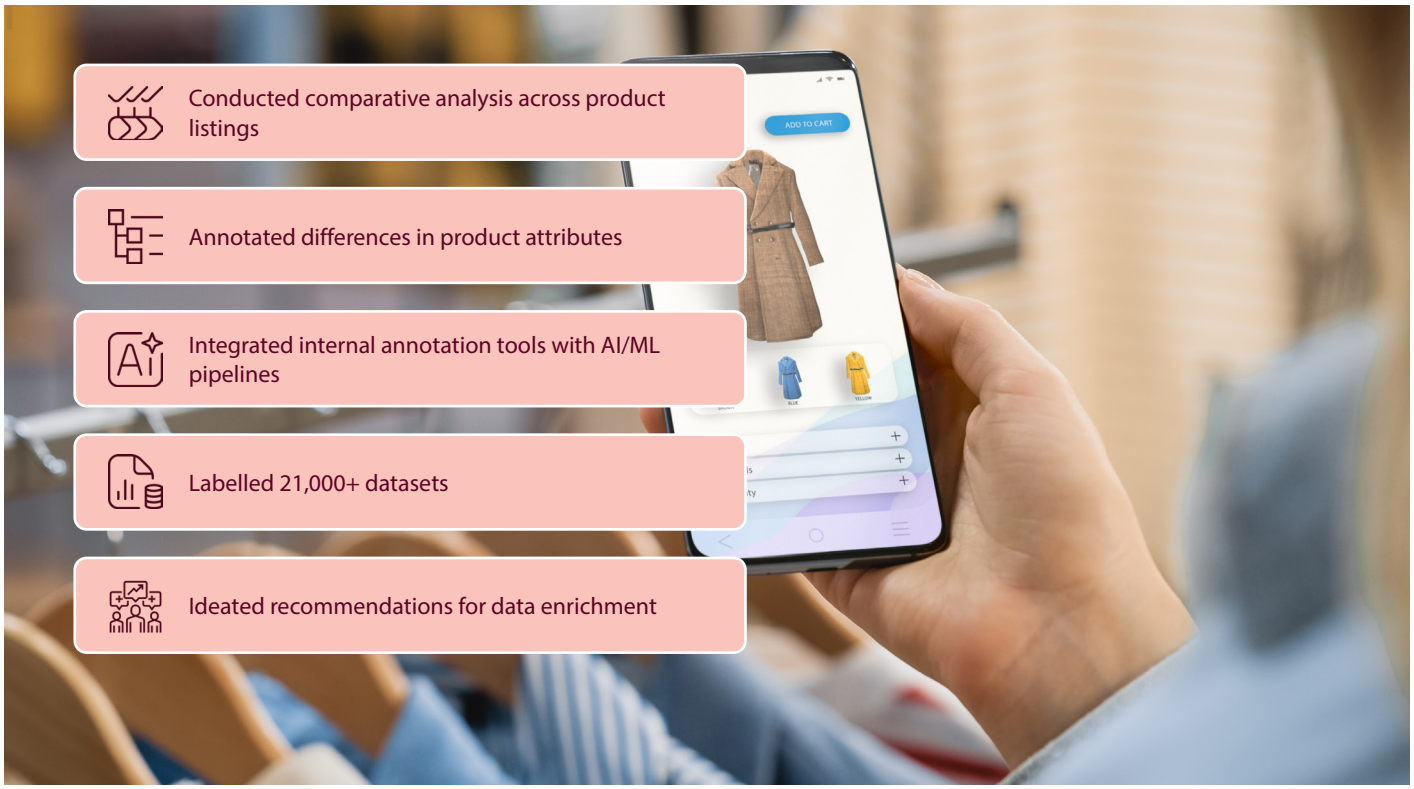
Knowing that the ML model's accuracy was entirely dependent on the quality of the training data, she could not afford to take any risks. She immediately began looking for a professional annotation and labelling vendor to support her in the project. Fortunately, she did not have to spend a lot of time in this search. The company already had an MSA signed with Infosys BPM, which Ashley decided to exercise for the project. After a meeting with the company's board, she signed on Infosys BPM for a 4-week association covering the annotation and labelling services. She immediately set up a meeting with Rakesh Rastogi, the Infosys BPM Project Lead, where she walked him through the existing process flow, detailed her requirements, and enlisted his support for effectively labelling the large volume of datasets.

Calling in for reinforcements

With a clear project scope at hand, Rakesh rounded up his team of data analysts and engineers to get things running. Together, the team spent some time studying the product listings and familiarising themselves with the large volume of datasets.

Noting the importance of staying accurate and leaving no room for ambiguity, Rakesh and the team then began with their detailed comparative analysis. They studied and compared product listings across competitor platforms, while zooming in on specific attributes to define the different product strategies. In order to ensure complete accuracy in the datasets for the ML model, Rakesh led the team in manually annotating the differences in product attributes (e.g., title, image, description, specifications). The team supported this with an internal annotation tool designed to integrate with the AI/ML pipeline. Throughout the project, Rakesh and team worked closely with Ashley to verify the annotations and maintain the highest level of accuracy and consistency.

Approach summary

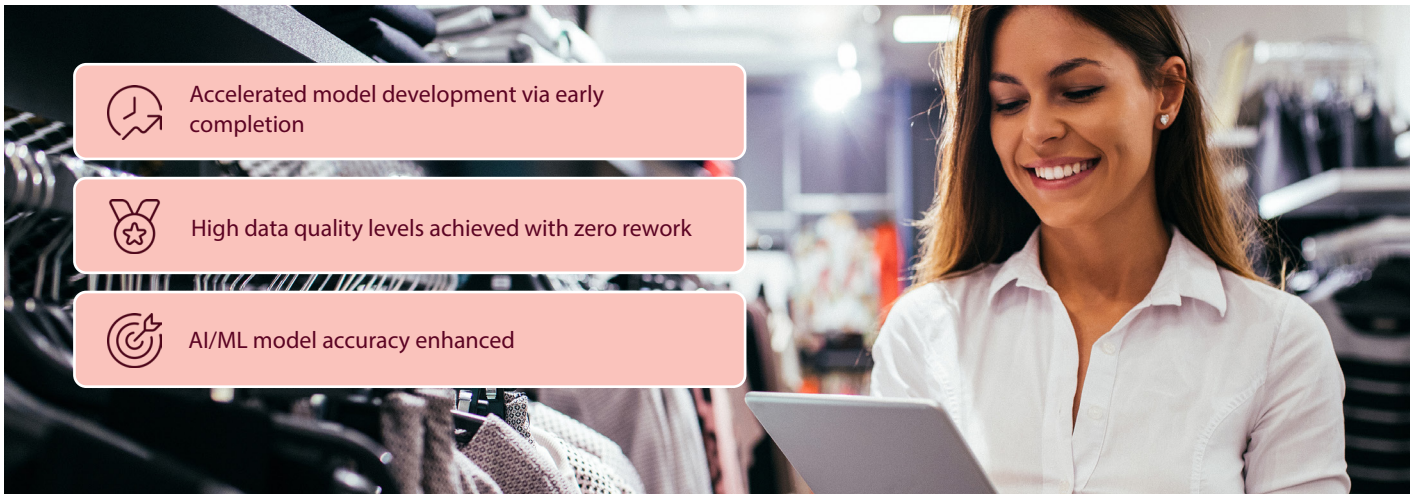


To Ashley's delight, they were able to successfully label over 21,000 datasets within just three weeks — a week ahead of the 4-week schedule. After delivering the labelled datasets, the team conducted elaborate ideation sessions, where they brainstormed new strategies for further enriching the datasets. The datasets were then used to train the AI/ML model, strengthening its long-term capacity.

A quick journey to optimal performance

Soon after Rakesh handed over the final annotated datasets, Ashley could see visible developments in the AI/ML model. The quality and accuracy of the annotations exceeded all expectations, with not a single instance requiring rework. This allowed Ashley and her team to feed the datasets fed directly into the ML pipeline, dramatically improving the efficiency of downstream model training and validation. The enhanced data quality soon reflected in significant performance improvements in the AI/ML model, enabling precise comparisons between product listings across platforms.

Key benefits



Moreover, the one-week early completion of the project added to the momentum. It allowed Ashley's team to accelerate the rest of the AI/ML development lifecycle, saving valuable time off their go-to-market roadmap. Overall, the comparative analysis with the annotated datasets gave Ashley fresh insights into competitor product strategies, enabling strategic business growth.

With the high-quality labelled datasets and the ML model performing at optimal efficiency, Ashley expressed her gratitude towards Rakesh and his team, applauding their consistent support throughout.

**Names have been altered to preserve the identities of the people involved.*

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next