

Seagate Accelerates Anomaly Detection Solution Scaling and Democratizes Machine Learning

Case Study



Company

Seagate Technology

Location

Normandale, MN

Contact

seagate.com

Industry

Data Storage

LYVELABS™
ISRAEL



Background

Seagate is the world's industry leader in mass data storage solutions. It has ownership and vertical integration of critical technologies in heads and media, and its wafer manufacturing requires advanced technologies. As such, it requires leading-edge analytics that enable it to address the complex problems and issues it faces.

Discover how Seagate's partnership with Vanti-Analytics enable a path to scaling an anomaly detection solution across 100+ operations, and democratize this process beyond the data science team.

Overview

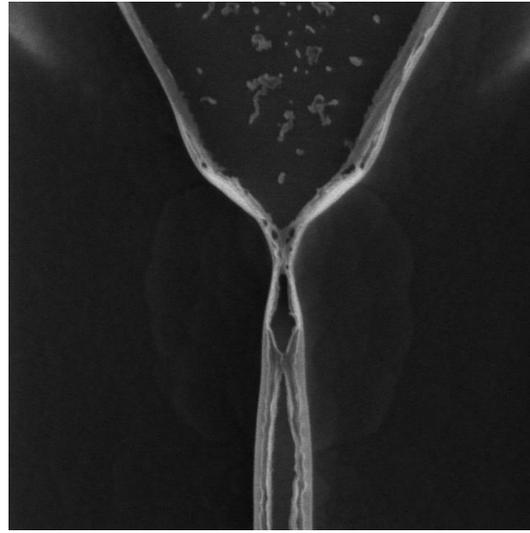
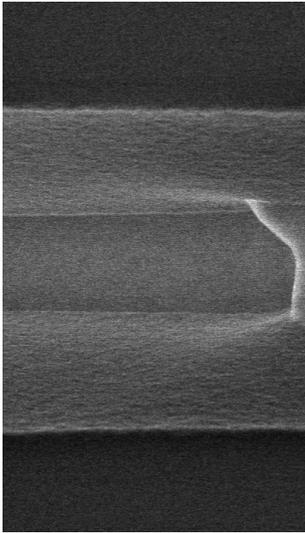
Multiple fault detection systems are currently in place throughout wafer manufacturing. Among their suite of solutions, there are image-based solutions that are trained to detect and classify known defects from the images that the factory tools produce. These were developed by Seagate's in-house data scientists and machine learning experts, and fully integrated to work with the existing factory systems.

Having conquered the area of known defects, the team wanted to address the next area of unknown defects; when even engineers cannot give them characteristics or predefine them. These defects are previously unseen and they may appear in any image, anytime.

Problem

The Global Wafer Systems team at Seagate had a very challenging task. There are hundreds of operations that generate various images, and within these sets of images, they may have very high variability. For each operation, there could be different fields of views for the same features. The solution they needed to come up with must be robust and fit for purpose, fully integratable with the existing factory systems, and, most challenging of all, scalable to other operations to achieve full coverage.

The initial solution that they developed was robust enough to deploy and integrate in the factory, but the solution required a lot of resources to label the data and perform the iterations required. The team already knew from experience that it was an unsustainable approach if they wanted to scale the solution to 100+ operations. In addition, because this problem is about unknown Seagate internal defects, humans labeling data becomes an exercise in judgment and defeats the purpose of data-driven anomaly detection.



Examples of metrology images for defect detection analysis

Lyve Labs Israel, Seagate's Innovation Center was opened in February 2020 in Tel Aviv, Israel. Lyve Labs' mission is to form partnerships with external innovation in the Israeli ecosystem in order to enable innovations by providing simple, secure, and efficient ways to work with exabytes of data. The initiative empowers the seamless movement of data, optimizing its business value both in flight and at rest. When they found cnvrg.io who offered an end-to-end machine learning platform to build and deploy AI models at scale, they thought their solution could be a good fit with Seagate's data scientists.

Solution

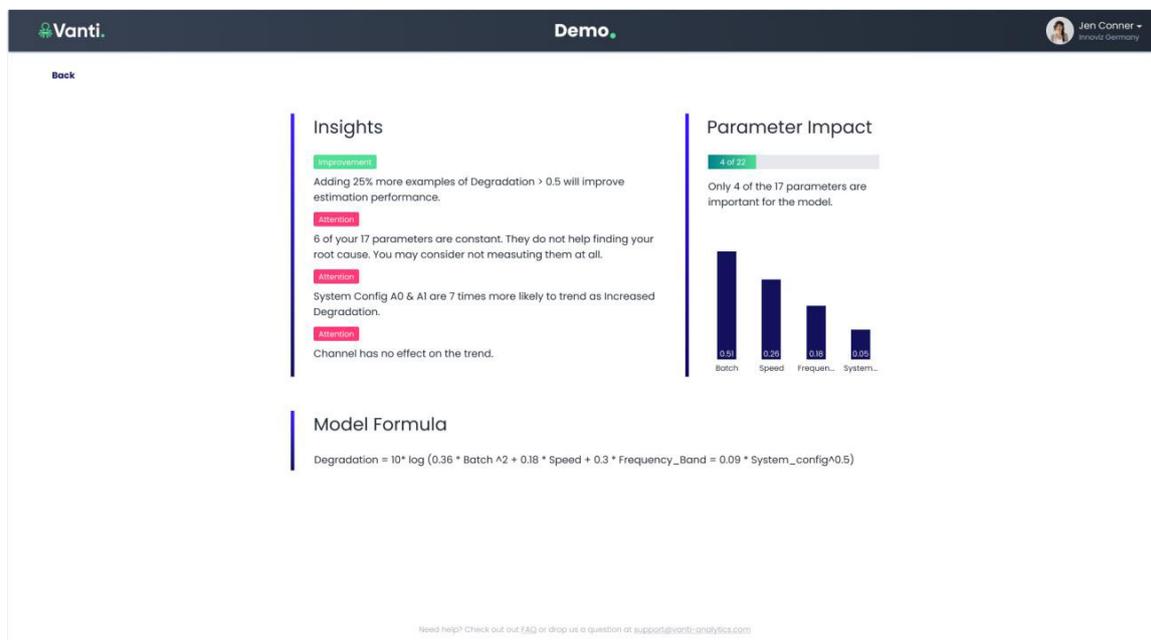
Seagate partnered with Vanti-Analytics to address this scaling problem. Vanti worked with Seagate on a proof-of-concept to demonstrate the applicability of Vanti's clustering algorithm. Vanti was able to demonstrate the advantages of their platform.

Vanti's self-service platform allows development and deployment of ML models with little to no data science knowledge required. Model development that would have taken the team days to do now took only a few hours. This is mostly because there are automated insights generated by Vanti's platform that include explanations about failures, clusters with higher likelihood of failures, and features that contribute to the model's prediction. These insights would reduce the time it takes a data scientist to develop and evaluate exploratory models.

In addition, Vanti's platform also provides a simple way for experts to add their contextual knowledge to the model, without the need for additional data science resources. Seagate Internal

Due to the complexity of the wafer manufacturing process, adding contextual information or domain knowledge into the process of model building is crucial to developing a solution for practical use in production. In one example, the insight generated by the platform indicated that there would be an improvement in the model accuracy if product code differences were considered and introduced in the training process. This immediately improved the model's performance by 18%.

(These screenshots are not from the Seagate use case.)



Vanti Demo

Back

Insights

- Improvement:** Adding 25% more examples of Degradation > 0.5 will improve estimation performance.
- Attention:** 6 of your 17 parameters are constant. They do not help finding your root cause. You may consider not measuring them at all.
- Attention:** System Config A0 & A1 are 7 times more likely to trend as Increased Degradation.
- Attention:** Channel has no effect on the trend.

Parameter Impact

4 of 22

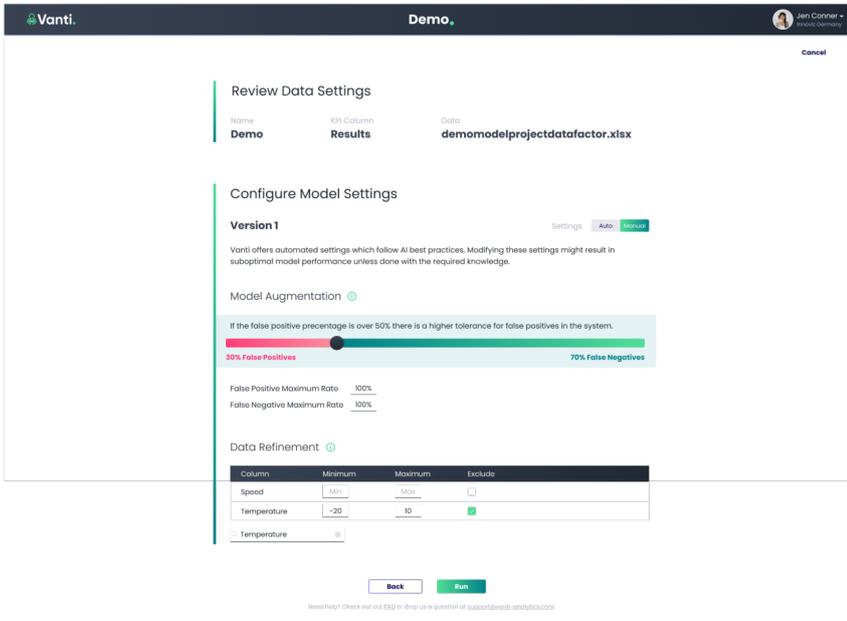
Only 4 of the 17 parameters are important for the model.

Parameter	Impact
batch	0.61
Speed	0.28
Frequen...	0.13
System...	0.02

Model Formula

$$\text{Degradation} = 10 \cdot \log(0.36 \cdot \text{Batch}^2 + 0.18 \cdot \text{Speed} + 0.3 \cdot \text{Frequency_Band} + 0.09 \cdot \text{System_config}^{0.5})$$

Need help? Check out our FAQ or drop us a question at support@vanti-analytics.com



Vanti's platform enables manufacturers to develop, deploy and maintain AI models at scale.

The self-service platform harnesses data directly from the shop floor to allow manufacturers to identify, prevent and explain defects in their production processes so they can achieve their quality and efficiency goals.

"We recognize the need to create a workspace where business experts and data scientists can work collaboratively, and Vanti's platform provides this, along with a specific use case where we can immediately deploy and harness the value of the work. Having already gone through the experience of deploying our models, we know what it takes to scale and operationalize solutions. But providing a platform for engineers for self-service machine learning will allow our data scientists to focus more on understanding the business problems well and ensuring that the data is accessible and good quality. The data scientists will still need to manage the model implementation, including setting best practice for data science standards, and ensuring that the models are fit-for-purpose, but the engineers can quickly inject their contextual knowledge into the models. It then creates a virtuous cycle for scaling solutions."

Agnes Zarate
Sr Manager of the Global Wafer Systems
Seagate

The Outcome

Utilizing Vanti's platform showed two distinct advantages.

First, the speed at which the iterations can be performed is remarkable. The generation of insights leads the data scientist or modeler to the next logical iteration to explore without having to experiment on various approaches. Model development that would normally take several days can now take just hours.

Second, the model development can be democratized. Instead of data scientists or machine learning engineers having to do the exploration, a subject matter expert can do the iterations quickly and introduce their contextual knowledge quite easily, bypassing the usual limitation on data science resources that may delay the development and integration of models. This will allow faster scaling and ownership of the solutions. This platform is not limited to a single algorithm with parameter tuning, but it can actually automatically pick the algorithm and then get insights from the data. In addition, it works with unsupervised learning, without the need for labeled data.