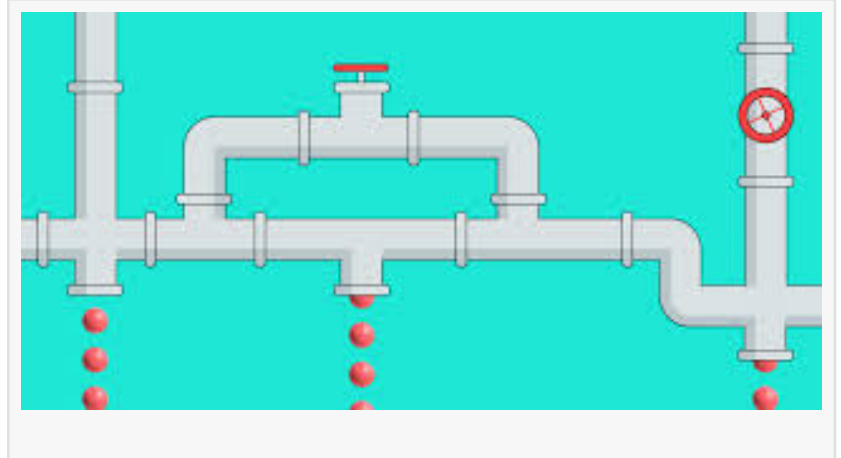# How Explorium Upgrades A Data Pipeline

SAN FRANSISCO, CALIFORNIA, USA, March 10, 2021 /EINPresswire.com/ -- How Explorium Upgrades A Data Pipeline

Let's use the example of a factory that makes computers. There is a need to have a steady pipeline of parts and raw materials. This necessity can be approached in two ways. The first way is to simply look at what was used during the last batch and make a new order every time. The second is to create a pipeline of steady suppliers, established channels, and a chain that works to deliver the same results every time.

Trying to do things from scratch every time might work once or twice, but it's not really scalable to any degree. It means focusing too much on the small things and not enough time spent on prioritizing results and impact. However, spending the time to build a pipeline, means automating a lot of the basic work and focusing on getting the most out of operations. Believe it or not, machine learning (ML) works the same way. The data used is essential, so building paths and pipelines that let you access it on demand is critical. Unfortunately, that's not always the case.

Even with the best-designed ML algorithms, time spent scrambling to find the right data to feed them every time they need to be retrained means wasting a lot of resources on gruntwork. That's where a data pipeline can benefit from Explorium's [data science platform](#).

What's in a data pipeline?
It's not inaccurate to say that ML models rely on data, but what we really should be saying is that to keep running ML models need reliable data pipelines. A model that uses one set of data once isn't really useful when there is a need to continuously make new predictions and gain fresh insights. Think about it this way — the world in which any model is running changes constantly, so why wouldn't the data?

However, simply finding entirely new data every time is a one-way ticket to poor results. No, what is needed is a data pipeline that is constantly adding new data to current sets without

requiring any upkeep. A data pipeline means that models keep running smoothly and remain relevant as new data emerges. With Explorium, it means the most relevant and up-to-date data on-demand.

Explorium and the data pipeline, reimagined
It's just not feasible to rebuild a data pipeline every time we need to rerun ML models, and Explorium makes sure that's never the case. However, Explorium goes well beyond simply having a place in the cloud to store datasets so that each new model can be easily run, but we're getting ahead of ourselves. Let's first look at how Explorium builds a data pipeline:

First, the data connects to Explorium. No matter where this data is kept — databases, apps, analytics tools — all of them can be connected to the platform. Once there, then the data pipeline is visible with all the relevant connections in a tree structure.

Next, run the data pipeline. With the right data connected, simply run the data pipeline to create a structure that works for the specific predictive question. Now there exists a persistent representation of data that is easily accessible any time there is a need to re-run models. Moreover, if the data changes, the pipeline will incorporate those changes when it is run again.

At the same time, Explorium adds thousands of external data sources to the pipeline. More importantly, though, it creates connections to each data source that gives continuous, real-time access.

Now that the pipeline is built, Explorium's real magic starts. It enriches data, trains and tests your models, and uses our automated feature engineering capabilities to generate impactful, relevant insights with these models. This can be done directly on Explorium or with their SDK and API to bring enriched data into your existing models.

It's really that easy. The best part of Explorium is being able to skip the grunt work and use expertise and domain knowledge to maximize ML models' potential.

Stop overthinking it, start building better pipelines

Explorium helps take the unnecessary steps out of ML and a data pipeline, meaning time to focus on results. Instead of having to rebuild a pipeline, data can be uploaded, the platform can be run, knowing that every time models need to be retrained with the most up-to-date versions of a dataset, it's all just a click away.

Explorium offers a first of its kind external data platform powered by an external data gallery and feature engineering. By automatically connecting to thousands of external data sources and leveraging machine learning to distill the most impactful signals, the Explorium platform empowers data scientists and business leaders to drive decision-making by eliminating the

barrier to acquire the right data and enabling superior predictive power.

Shelby Blitz
Explorium
+972 58-7333765
email us here

This press release can be viewed online at: https://www.einpresswire.com/article/536615597