



CASE STUDY

Actium Produces Breakthrough ML Models by Accelerating R&D with Prefect Cloud

Actium helps healthcare organizations coordinate every level of their health system by delivering data insights and orchestrating Next Best Actions.

COMPANY



INDUSTRY

Hospital + Healthcare

HEADQUARTERS

Palo Alto, CA

PRODUCTS USED

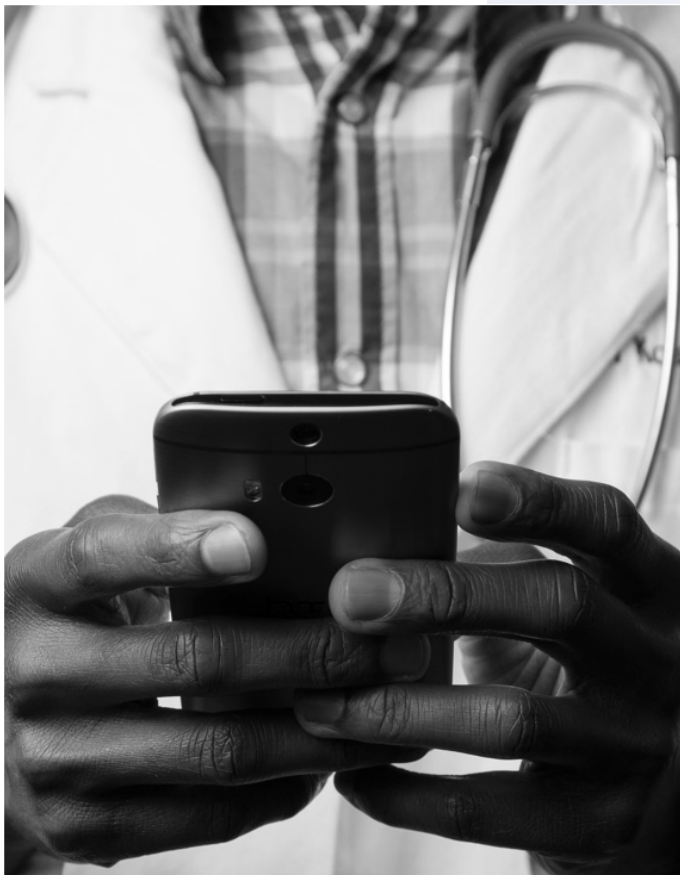


99%

Decrease in implementation time
for ML Model Training

350

Experiments run at one time



SYNOPSIS

Introduction

Actium helps healthcare organizations by delivering data insights on operational objectives and determining Next Best Actions for patients.

Challenge

Rapidly expand data science work to train ML models for production-ready data pipelines to modernize predictive patient treatment.

Solution

Orchestrate with Prefect to automate model training and manage R&D experiments at scale, allowing the Actium team to develop production-ready pipelines rapidly.

Results

Prefect's workflow semantics and centralized platform streamlines Actium's dataflow, accelerating R&D into high-performing predictive ML models and automating ongoing retraining.

Introduction

Driving innovation in healthcare, Actium helps healthcare organizations coordinate every level of their health system by delivering data insights and orchestrating Next Best Actions. Powered by comprehensive analytics, the HealthOS platform provides personalized patient treatment recommendations processed by machine learning algorithms. By providing organizations with Next Best Actions at an organizational and individual level, Actium's HealthOS platform provides an intelligence layer to health systems to enable data-driven patient care and dynamic company activity.

Challenge

To fully deliver on Actium's vision of enabling health systems with a broad catalog of Next Best Actions (NBAs) for patients, the data team needed to take on a massive expansion of the data science work, as the data coverage included everything from cancer screenings to orthopedics to cardiology. Facing significant time pressure to stay ahead of competitors, CTO Joe Schmid and his team ramped up R&D around NBA machine learning models by experimenting on large volumes of healthcare data.

" We had outstanding data scientists from top institutions like Carnegie Mellon, Stanford, etc. but we needed to give them the tools and infrastructure to successfully meet our challenges.

JOE SCHMID, CTO

Speed and scalability were paramount, and the team needed lightning fast execution and the ability to rapidly scale out by running hundreds of experiments in parallel. The initial infrastructure led the team to Dask, a parallel computing library that natively scales Python, to unlock scalable data science and meet their specifications.

" Our initial experience with Dask for loading large datasets, engineering features, and training models in parallel on a cluster was incredibly promising. At that point, we knew Dask would likely be an important piece of the puzzle.

JOE SCHMID, CTO

However Dask doesn't provide higher-level workflow abstractions, which could enable the team to orchestrate and automate workflow processes. The next challenge was to create

data science pipelines as parameterized workflows that covered all aspects of running machine learning experiments.

" As we started to look at the initial Prefect Core open source release in Q2 of 2019 we got incredibly excited about the possibility to combine the cluster computing capabilities of Dask with the workflow semantics of Prefect.

JOE SCHMID, CTO

After a brief proof of concept automating data extraction from internal warehouses, Joe and his team quickly began orchestrating parallelized pipelines with Prefect Core on a 100-node Dask cluster and saw promising early results. By building their cluster with AWS on Kubernetes, they enabled rapid up-and-down scalability and could begin training machine learning models.

Solution

The early results of running Prefect Core on a 100-node cluster exceeded expectations, accelerating machine learning research output and reducing team bandwidth. The team developed pipelines utilizing Prefect Mapping with Dask Executors for parallel execution, balancing workloads with Task Resource Tagging to delegate resource-intensive tasks to high-memory Dask workers. The newfound efficiency allowed data scientists to iterate various machine learning approaches by easily running hundreds of experiments very quickly and reliably.

" I used the parallelized hyperparameter tuning (incredible) to run about 350 experiments in 30 minutes, which normally would have taken 2 days.

ANDREW WATERMAN, DATA SCIENTIST

The breakthroughs in the R&D phase provided by Prefect Core's workflow semantics naturally lent themselves to Prefect Cloud, a centralized platform for the team to monitor flows and logs on distributed Dask workers. With the added visibility into ongoing processes and the ability to execute workflows in remote environments, Prefect Cloud's built-in feature set provided essential infrastructure for the team to confidently push to production. Additionally, Cloud's Task Concurrency Limits prevented overwhelming data stores when pipelines needed to store results from hundreds of experiments in a database. The Cloud UI also provided monitoring of active Prefect Agents,

whose health is integral to executing workflows on Actium's own Kubernetes infrastructure.

Prefect's Hybrid Execution Model assured the team their data-sensitive development was secure, using a remote Dask environment to run their code on existing infrastructure, empowering rapid experimentation and reducing R&D implementation time.

"...as a vendor to covered entities under HIPAA, security and compliance are huge issues for us. Prefect's hybrid execution model was key to being able to run with Cloud while none of our client data left our infrastructure."

JOE SCHMID, CTO

Expediting the machine learning modeling allowed the team to prioritize their rapid development, eventually leading to the production of many successful models now used in production. The team's breakthroughs were immediately available to patients, developing models recommending various service lines such as orthopedics, oncology, and cardiology.

Results

Reduced implementation time for ML model training.

Prefect allowed data scientists to easily run hundreds of experiments very quickly and reliably using Dask parallelism to iterate through various machine learning approaches under time-sensitive conditions.

"By cutting our machine learning experiment time from 2 days to 30 minutes we had a huge gain in the number of experiments we could run each day. By allowing us to iterate faster, we could quickly experiment with more machine learning approaches, which ultimately got us to production ready models very quickly."

JOE SCHMID, CTO

Rapid R&D phase yielded breakthroughs.

Breakthroughs in successful ML modeling produced many

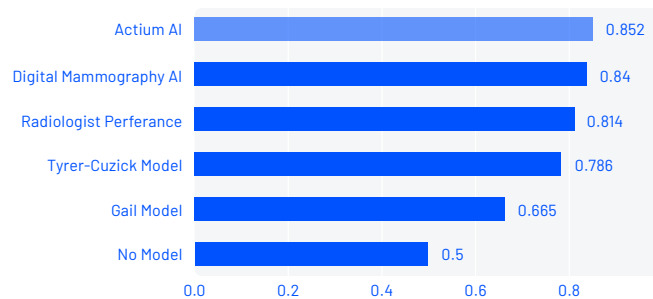
models used in production today, recommending patients for various service lines such as orthopedics, oncology, and cardiology.

Notably, one of the team's latest breakthroughs is achieving the highest AUC score in Breast Cancer Detection performance across models, surpassing the Gail Model and Tyrer-Cuzick Model scores.

"Most importantly, this infrastructure was critical to achieving our objectives: our R&D was successful and we now had a large number of machine learning models ready to be used in production."

JOE SCHMID, CTO

Breast Cancer Detection Performance Across Models



AUC Score: Measures overall classification performance, higher is better.

5-15x

Patients identified by the model were 5-15x more likely to have breast cancer.

675k

Patients scored by the model. This approach is scalable, not limited by surveys & image data.

Hybrid Model enabled seamless compliance process.

Sensitive data and code never passes through Prefect Cloud's infrastructure, empowering the team to quickly overcome compliance hurdles and expedite R&D.

"...as a vendor to covered entities under HIPAA, security and compliance are huge issues for us. Prefect's hybrid execution model was key to being able to run with Cloud while none of our client data left our infrastructure."

JOE SCHMID, CTO

Data team regains valuable time by implementing both Prefect Core and Prefect Clouds.

Prefect Cloud enhanced visibility and logging into long running flows and distributed tasks is safeguarded by failure alerts via Slack, allowing data scientists to spend their time and energy on more important projects than the constant monitoring of fragile jobs.

“ We now build all of our data pipelines using Prefect. Data scientists and engineers find it easy to work with Prefect as it lets them focus on what the flow needs to do rather than error handling, retry logic, logging, etc.

JOE SCHMID, CTO

The team streamlined imminent ML development with Prefect flows that automate ongoing predictive model updates on a nightly basis and model retraining on a weekly or monthly basis.

Data scientists and engineers continue to develop Prefect flows for data pipeline needs, as it allows them to focus on projects at hand, rather than error handling, retry logic, logging, etc.

“ Homegrown data science pipelines are often cobbled together and lack repeatability. By embodying full data science pipelines as Prefect flows, we know that they will run reliably which means we don't have to worry about experimental runs failing often.

JOE SCHMID, CTO

Ongoing Efforts

Prefect's utility allows streamlining for future data pipelines.

Ease of use allows Actium data scientists to expand usage of Prefect beyond ML to production-grade data engineering, standardizing future implementations of data pipelines.

“ While we initially focused on using Prefect for our data science pipelines, we now use it for data engineering as well. Specifically, we built an ETL pipeline to reliably transfer data from our production database to our data lake and transform it prior to our machine learning pipelines. Prefect Cloud's scheduling and parameterization allows us to easily run the flow nightly or on an ad hoc basis with customized lists of table names and selection criteria.

JIE LOU, DATA SCIENTIST

“ Prefect is very beginner friendly: it took me five minutes to write and execute my first simple Flow with @task. Watching it execute with flow.run() was like magic!

JIE LOU, DATA SCIENTIST