tecton + vital

Vital Speeds Up Delivery of Machine Learning Products by Genericizing Feature Code & Improving Testing Infrastructure

> *"Our previous process involved computing features separately at training and inference time. With Tecton, not only do we have confidence in our computations due to the testing infrastructure, but our engineering and data science teams can now leverage the same code for all our use cases, online or offline. By genericizing our feature code, we have significantly sped up our ability to deliver models."*
>
> — Felix Brann, Head of Data Science at Vital

OVERVIEW

About:

Vital is a healthcare software company whose mission is to transform the care experience for patients, clinicians, and staff. To do this, Vital engages patients throughout emergency department (ED) and inpatient visits, driving improvements in both clinician efficiency and patient satisfaction.

Challenge:

Vital builds healthcare software to improve the overall patient care experience. To develop their predictive healthcare products more efficiently, Vital's Engineering and Data Science teams needed a solution to improve their machine learning featurization infrastructure and standardize batch, streaming, and real-time data transformation pipelines for use across predictive products.

Solution:

It took Vital 3-4 months to create the infrastructure that allowed them to deploy models more quickly using their generic feature code. Tecton's declarative framework has enabled the team to design machine learning features as code and rely on Tecton to compile, orchestrate, and maintain the associated data pipelines. Along with reducing the testing burden and the risk of training / serving skew, Tecton enables them to re-use feature code across use cases. Today, Vital has 3 Tecton-powered models in production and plans on additional model deployments in Q4 2022.

Results:

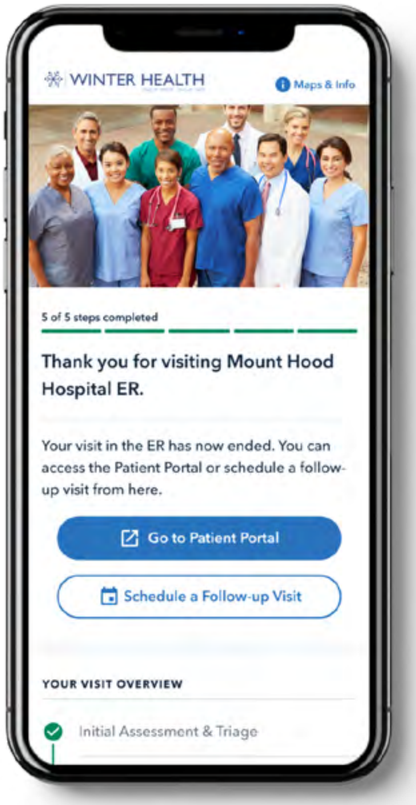
Vital successfully integrated Tecton into their infrastructure in under 4 months so their teams can now:

- **Genericize feature code and speed up delivery of future ML products** by using Tecton's SDK to build and manage features as Python files
- **Improve their testing infrastructure** by using Tecton to generate consistent data offline for training and online for inference
- **Leverage real-time and streaming data** to build and compute on-demand features with data only available at request time
- **Automate ongoing feature materializations** to ensure that the system recomputes feature views as data sources change

Discover how Vital's data science and engineering teams leverage Tecton to genericize feature code and improve testing infrastructure to build and deploy predictive products in record time.

Tecton and Vital, in practice

With Tecton, Vital powers its customer-facing emergency department (ED) predictive product with request-time input signals and solves the "cold-start problem" by building machine learning feature feedback loops.



5 of 5 steps completed

Thank you for visiting Mount Hood Hospital ER.

Your visit in the ER has now ended. You can access the Patient Portal or schedule a follow-up visit from here.

YOUR VISIT OVERVIEW

Initial Assessment & Triage

Vital's flagship product, ERAdvisor, gives patients up-to-date, real-time information about their ED visit. Patients typically experience ED visits in six distinct stages: registration, triage, bed assignment, waiting for a provider, waiting for a clinical decision, and either an admit or discharge. Vital builds models that predict wait times associated with these steps and uses artificial intelligence, machine learning, and natural language processing to build beautiful, functional software that requires no training and is easy to implement and use.

ERAdvisor provides patients with real-time, personalized updates in the ED. By combining EHR data, artificial intelligence, natural language processing, and a sleek user experience, patients can view and share information about their ED visit—all via a web app accessible from any handheld device or computer. The app updates patients throughout their ED visit with step-by-step automatic progression, from arrival to admission, discharge, or transfer. As a patient moves through the ED, the application predicts how long they will wait at each step of their journey.

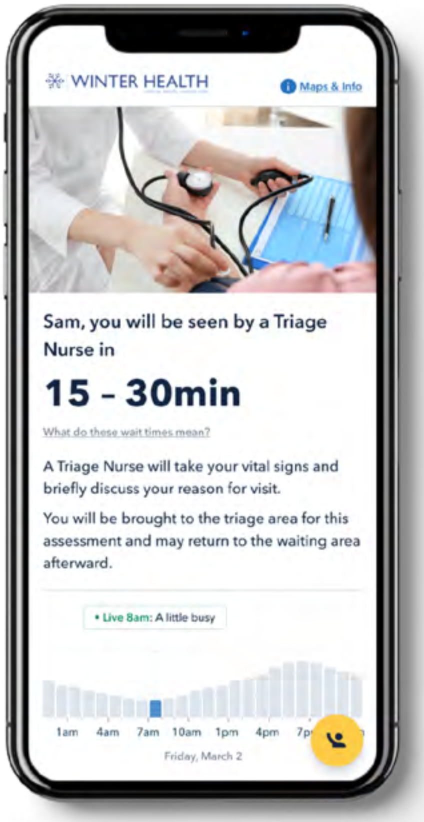
Challenge

Vital deploys ERAdvisor in new facilities every month, and their clients expect the application to be up and running as soon as data is available. To make predictions on demand that are customized to each patient's care journey, ERAdvisor's predictive model uses freshly computed features.

These features range from simple current time of day, to more complex number of patients in the queue ahead, to very sophisticated derived aggregated rolling window features. Each facility they deploy into has very different characteristics for these features; for example, the average number of patients at a busy facility may be 5x that of a quiet one.

Tecton helped Vital scale ERAdvisor across facilities by tackling three key challenges:

- **Cold-start problem:** By designing and running feature feedback loops with Tecton, ERAdvisor can provide correct wait-time predictions despite having very little data to power its model in newly deployed facilities.
- **Near real-time data:** With Tecton's feature platform, Vital abstracts the complexity of integrating real-time signals for online inference.
- **Time to production:** By standardizing its feature compute and serving layer while maintaining real-time feature computations, Vital can quickly deploy ERAdvisor in new facilities without starting from scratch every time.



Sam, you will be seen by a Triage Nurse in

15 – 30min

What do these wait times mean?

A Triage Nurse will take your vital signs and briefly discuss your reason for visit.

You will be brought to the triage area for this assessment and may return to the waiting area afterward.

• Live 8am: A little busy

When launching at a new facility, ERAdvisor needs to be able to accurately predict wait times despite having seen very little data at that location. In order to do this, it learns dynamics from Vital's other facilities and normalizes the computed features in order to apply those dynamics at the new facility.

Solution

To solve the cold start problem, ERAdvisor uses information collected at all Vital facilities to learn shared dynamics between features and predictions. ERAdvisor uses a complex feature feedback loop that treats aggregated feature information from a rolling window distribution as raw input data for new feature computations. Instead of predicting the number of minutes that an individual patient will wait for a bed or before seeing a doctor, the system predicts how much longer they will wait than what is recently normal for that facility.

To generate these normalized features and automate the feedback loop, Vital's Engineering and Data Science teams use Tecton to automate the pipelines that ingest computed features and convert them into summary statistics. The model uses purpose-level aggregated feature summaries from a rolling window to convert absolute numbers into deviations from the baseline for that facility.

Furthermore, some of Vital's more sophisticated pipelines ingest data only available at request time when a patient is refreshing the Vital wait-time prediction. In such cases, Vital pulls only the latest, freshly computed feature values from Tecton's online store to make predictions. Vital currently has 15 event streams running through Tecton.

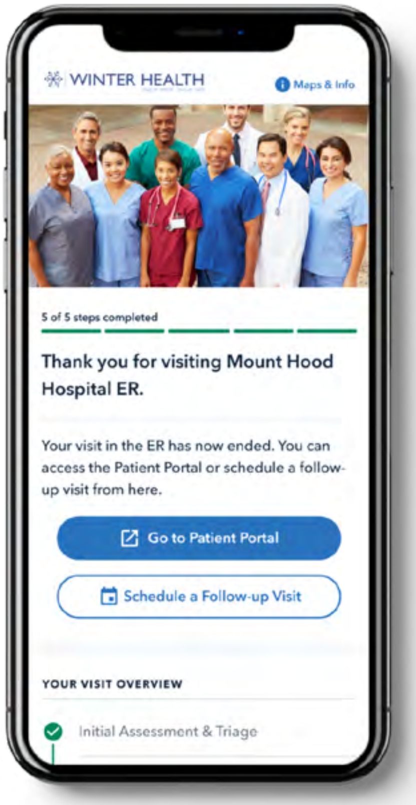
Results

With the help of Tecton's feature platform for machine learning, Vital develops, productionizes, operates, and re-uses feature pipelines that are constantly computing up-to-date feature signals used for near real-time inference across ERAdvisor ED facility deployments.

> *"At Vital, we work hard to improve ED care experiences by building smart applications for patients and their families. Our ERAdvisor product needs live features to provide patients with wait time predictions, and setting up a robust feature serving pipeline was a big deployment challenge. By adopting Tecton into our stack, we were able to build, in just under six months, a system that computes on-demand features with data only available at request time. Now that our teams have developed expertise in Tecton to build and automate real-time ML features, we're excited to keep building models which impact patient lives for the better."*
>
> — Felix Brann, Head of Data Science at Vital

With Tecton, the team has been able to produce models that predict wait times with high accuracy as soon as data starts flowing from a new facility. Moreover, the rolling window aggregations allow the model to automatically respond to regime changes in the ED environment. Vital can now roll out similar wait time models (which would otherwise take months to develop) in 2-4 weeks.

Vital plans to continue using Tecton with Databricks to build continuous-mode streaming pipelines (feature recomputation within seconds after Tecton reads an event) to power future machine learning products.



Thank you for visiting Mount Hood Hospital ER.

Your visit in the ER has now ended. You can access the Patient Portal or schedule a follow-up visit from here.