

# Travtus – Amazon SageMaker Machine Learning Project

## An Ubertas Consulting Case Study

### The Challenge

ADAM automates contact centre and operations related to property maintenance. The product uses machine learning to automate multi-family property management, by first reviewing resident communications and service requests to gather intelligence on the property, then cross-referencing this information with renewal data and other key property metrics; this provides a 360-degree view of the property.

Data preparation is key. The process of preparing raw data, so that it is suitable for further processing and analysis, includes collecting, cleaning, and labelling raw data into a form suitable for machine learning (ML) algorithms, before exploring and visualizing the data. Data preparation consumes up to 80% of the time spent on an ML project. Using specialized data preparation tools is an important and effective way to optimize this process at Travtus.

The current architecture utilises EC2, Lambda and S3, and the MLOps approach is highly manual and time-consuming. Travtus were keen to explore Amazon SageMaker to determine its strengths and capabilities, as well as the activity and cost of migration and modernising to this Service.

As part of the program of works several pain points in the current state architecture were identified. These areas of concern served as the basis for the design and scope of a proof-of-concept project.

### The Solution

Amazon SageMaker features make it easier for developers to automate and scale all steps of the end-to-end machine learning workflow. This is enabled through powerful capabilities, such as: faster data preparation; a purpose-built repository for prepared data; workflow automation; greater transparency of training data (to mitigate bias and explain predictions); distributed training capabilities to train large models, up to two times faster (than would otherwise be possible with today's machine learning processors), and model monitoring on edge devices.

With this in mind, and considering the challenges identified and requirements gathered, a proof-of-concept (PoC) project was agreed with the goal of demonstrating the value of SageMaker to Travtus, and how it could help solve key problems for the business.



# Travtus +

### About Travtus

Travtus is a London based Artificial Intelligence Research & Development company. With their long-standing experience in Real Estate Operations and with the latest research in Deep Learning and Conversational Interface design, they strive to bring automation to Real Estate and are the creators of a leading Artificial Intelligence platform (ADAM), which enables Real Estate companies to provide exceptional customer service.

ADAM helps to bring efficiency and automation across all areas of Property Management, by engaging with customers in the way they prefer. From voice and text, to web chat and apps, Adam helps customers access the services they need 24 x 7 x 365.

---

“ We have an increasingly complex ML stack and a lot of models which run in Production, as a business, [Sagemaker will give] us more control, more structure, more predictability. It will also help us manage quality better as well as improving practices and development speed. [The PoC] helped provide structure & flexibility long term and also [provides] control over production workloads. ”

**Andrew Day**  
**Co-Founder at Travtus**

---

The scope of the project was a pre-trained PyTorch text Sentiment extraction model, with SageMaker pipelines. The objective was to demonstrate how to execute model training and development workflow using Amazon SageMaker.

The method employed was to use a toy problem of fine-tuning a pre-trained Pytorch text clarification model on a data set provided by Travtus. The trained model would be hosted as a serverless service, which was built into Lambda execution functions in Service Accounts, as part of an overall Travtus Model Service.

This then showed how SageMaker pipelines brings CI/CD best practices into Machine Learning. This feature was demonstrated through a SageMaker Studio Notebook, for ease of use by Travtus Data Scientists.

The second part of the project was to alter the pipeline to a pre-trained model – again, trained on a Travtus data set. This trained model was hosted as a real-time endpoint, which was then built into Lambda execution functions, in Travtus Service Accounts as part of the Model Service.

Travtus’ current architecture utilises EC2, Lambda and S3. Travtus were keen to explore SageMaker to understand its feasibility, and quantify costs, time and complexity benefits that can be leveraged from migrating to, and modernising on, the service.

Based on the challenges defined at the start of the project, the following outcomes were delivered:

- Executed a model training and deployment workflow using SageMaker
- Created a PyTorch pipeline, built to fine-tune a pre-trained RoBERTa model on custom data.
- Implemented CI/CD best practices into machine learning by utilising SageMaker pipelines
- Utilised advanced SageMaker capabilities
- Provided technical documentation and knowledge share.

## The Benefits

Amazon SageMaker proved to be the right modernisation path for Travtus. During the proof-of concept project we were able to solve all the agreed problem areas, and delivered the immediate benefits:

- Moving away from using EC2 instances removed memory errors and the overhead of managing long running jobs
- Data science teams no longer need to rely on DevOps teams to provide code and CI/CD changes to run machine learning models
- Model and data versioning no longer rely on time-consuming manual processes and the risk of introducing human error
- SageMaker brings a range of process related improvements to the team: new model training processes, robust lineage tracking and model reproducibility, improved production model monitoring
- Automated Model Retrains, to perform a new model training runs at scheduled intervals in time or once a certain threshold of new labelled data has been reached
- using Amazon Eventbridge Scheduler with SageMaker Pipelines
- SageMaker provides a scalable and flexible platform that allows Travtus to build high-performing models while still being cost-effective
- SageMaker Model Monitor allows users to monitor the deployed real-time endpoints.

Amazon SageMaker delivered the following new capabilities:

### S3 Data Capture

Model Monitor provides a Data Capture configuration to the real-time endpoint, which can ingest input data to the hosted model and prediction outputs. The data is stored on S3, in JSON Lines format.

### Data Drift

Model Monitor allows data scientists to analyse and monitor the data from the hosted mode. Monitoring jobs can be scheduled to run custom scripts on the captured data. These jobs can be used to detect data drift in the environment where the model is hosted.

## Why Ubertas Consulting

Ubertas Consulting is a Cloud consultancy specializing in Amazon Web Services. As an Advanced Focus Partner, AWS Migration Partner, AWS Channel Reseller (Solution Provider Program) and Well-Architected Framework Program Partner our mission is to transform the way organisations modernise using AWS. We assist companies drive innovation and build new capabilities through embracing “Cloud Native” technologies and modernising with Amazon Web Services (AWS).

To find out more about how Ubertas Consulting can help you take your next steps to migrating to AWS with confidence visit us at : [www.ubertasconsulting.com](http://www.ubertasconsulting.com) or contact us at [info@ubertasconsulting.com](mailto:info@ubertasconsulting.com)

We're here to help.