

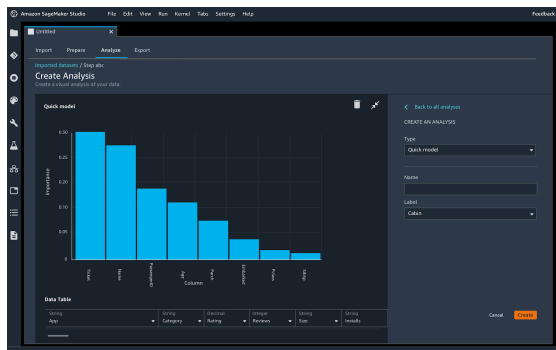
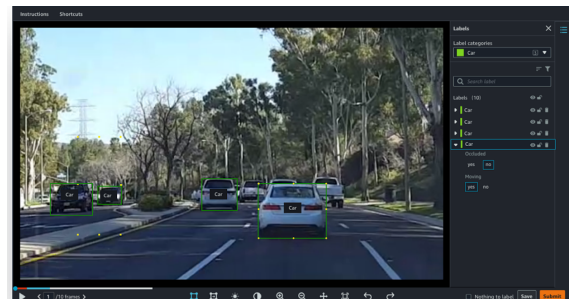
Amazon SageMaker

Introduction

Amazon SageMaker is a cloud based machine-learning platform that enables developers to create, train, and deploy machine-learning models in the cloud. SageMaker provides a variety of services at various levels of abstraction. At the highest level, SageMaker can be broken down into a series of cohesive components that users can leverage for their workloads.

Data Preparation

The first segment of services provide functionality for data labeling, data preparation, processing, and iterative experimentation. SageMaker Ground Truth is a managed data labeling service and interface that integrates and speeds up the process for manual training data labelling.

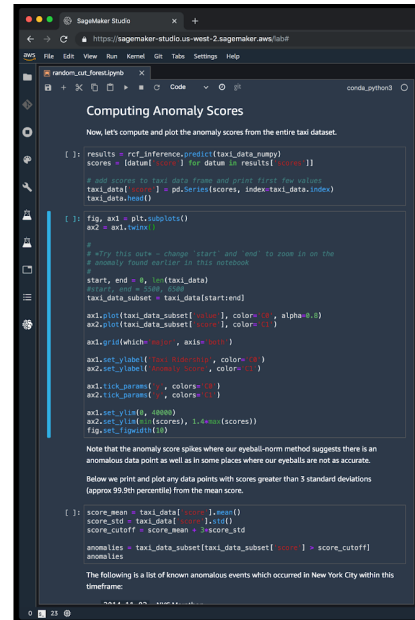


Once a dataset has been prepared, the data can be processed using SageMaker Data Wrangler. Data Wrangler allows the user to quickly normalize, transform, and combine features from the Data Wrangler interface. The interface provides interactive tools for data visualization and feature engineering to help in the process. Of course, custom data transforms can be written and imported.

Experimentation

While SageMaker provides a collection of built-in optimized algorithms and models, the tools for incorporating or developing models is extensive. One important aspect of model development is iterative development.

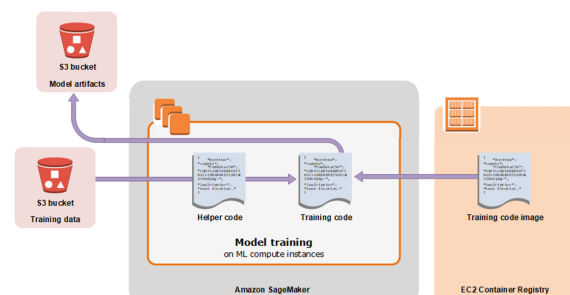
An extremely useful tool for experimentation is SageMaker Studio, a web based service to view, run, and share Jupyter notebooks operating on scalable resources. The notebooks are fully managed and can be secured with SSO, providing a tool for easy collaboration. This environment is perfect for playing around with the data, developing model architecture, debugging, optimization, and evaluating results.



Alternatively, AWS provides an AutoML solution, SageMaker Autopilot, to automatically build, train, and tune candidate ML models based on provided data. While limited to tabular data, Autopilot provides automatic data cleaning, preprocessing, algorithm selection, hyperparameter optimization, and distributed training. For those with little ML experience, this can get a decent solution off the ground quite quickly.

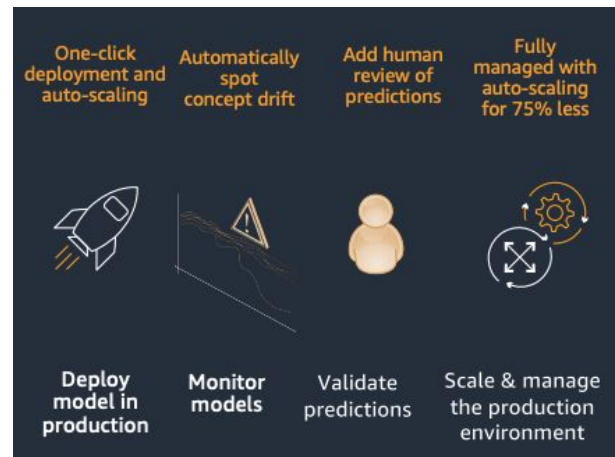
Training

Beside all of the prepackaged ways to train models on SageMaker, training can be quite flexible. A training job requires the user to specify an S3 bucket to locate training data, an S3 bucket to store the output, an ECR path to the inference training image, and to specify the requested compute resources. As mentioned previously, the inference image may contain a custom algorithm or an algorithm provided by SageMaker / AWS Marketplace.



Deployment & Management

After models have been built and trained, SageMaker can deploy models in two primary ways. First is by leveraging SageMaker model hosting services. SageMaker provides an HTTPS endpoint where the model is available for inferences. The endpoint can be configured to serve multiple models, multiple variants of a model, or combinations of models and variants. The endpoint can be configured to elastically scale the deployed compute instances.



Second, inferences across entire datasets are often desired. If a persistent endpoint is not required, SageMaker Batch Transform can be used to generate and index inferences for offline prediction.

CI/CD

Finally, to help manage machine learning workflows, SageMaker Pipelines can be used to trigger retraining, and redeployment. A Sagemaker pipeline can be triggered by newly available data via CloudWatch events, or upon new code check-ins. The data will be automatically processed, used to train the models, and automatically validated. Based on the validation, the models can be conditionally deployed. SageMaker Pipeline allows users to track numerous model artifacts, catalog model versions, metrics, and approvals for deployment.

Conclusion

AWS Sagemaker provides a powerful suite of tools for all steps of the machine learning workflow. This summary only scratches the surface of the functionality that I have found most useful. I encourage the reader to explore AWS Sagemaker, or at least keep their eye on its development.

Sources

<https://aws.amazon.com/sagemaker/>
<https://docs.aws.amazon.com/sagemaker>
<https://github.com/aws/amazon-sagemaker-example>