



Del notebook al pipeline sin dolor con AWS

Sebastián Sandoval García

Objetivo

Generar una flujo automatizado que permita predecir si un pasajero sobrevivió al hundimiento del Titanic.

Tomaremos como punto de partida un notebook tradicional de entrenamiento y lo convertiremos en un flujo de Amazon SageMaker Pipelines.

Objetivo



Herramientas

Amazon SageMaker



Amazon S3



Amazon S3 (Simple Storage Service)

- Almacenamiento de objetos accesibles desde la web (a través de API o HTTPS).
- Integración sencilla con otros servicios de AWS.
- Capacidades de seguridad, cumplimiento y auditoria
- Casos de uso: **lagos de datos**, sitios web, aplicaciones móviles, respaldo y restauración, archivo, aplicaciones empresariales, dispositivos IoT y **analítica de datos**.



Amazon SageMaker Studio

SAGEMAKER STUDIO

Prepare
data

Store
features

Detect
bias

Build with
notebooks

Train
models

Tune
parameters

Deploy in
production

Explain
predictions

Manage
and monitor

Amazon SageMaker Studio es un entorno de desarrollo integrado (IDE) basado en la web para preparar datos, crear, entrenar, implementar y monitorear sus modelos de aprendizaje automático.

Amazon SageMaker Studio

Computing Anomaly Scores

```
Now, let's compute and plot the anomaly scores from the entire taxi dataset.
```

```
[3]: results = ref[~ref['is_anomaly']] # remove anomalies
scores = [dat[0] for dat in data[['start', 'end', 'score']]]
taxi_data['score'] = pd.Series(scores, index=taxi_data.index)
taxi_data.head()
```

```
[4]: fig, ax1 = plt.subplots()
ax2 = ax1.twinx()

# Try to set date as 'start' and 'end' to zoom in on the
# anomaly found earlier in this notebook
start, end = pd.Timestamp(taxi_data['start'].min()), pd.Timestamp(taxi_data['end'].max())
taxi_data_subset = taxi_data[start:end]

ax1.plot(taxi_data_subset['value'], color='blue', alpha=0.5)
ax2.plot(taxi_data_subset['score'], color='red')

ax1.set_ylabel('taxi', color='blue')
ax2.set_ylabel('anomaly score', color='red')

ax1.tick_params(labelcolor='blue')
ax2.tick_params(labelcolor='red')

ax1.set_xlim([start, end])
ax2.set_xlim([start, end])
fig.tight_layout()
```

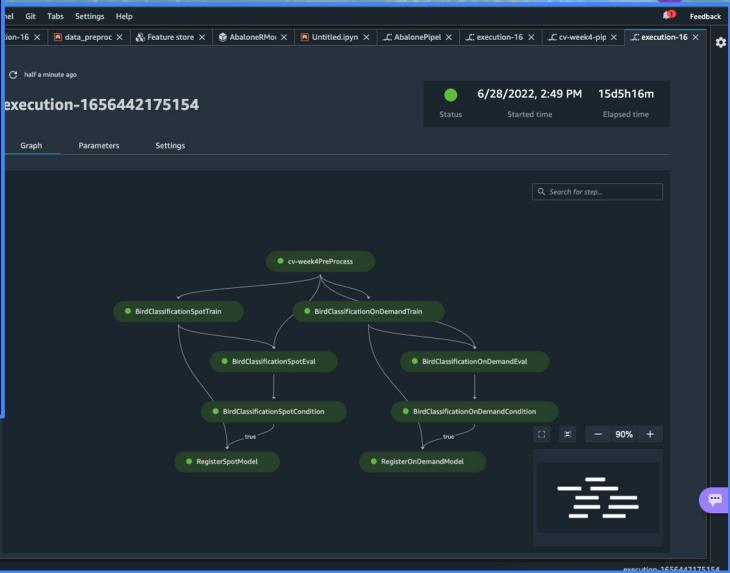
Note that the anomaly score spikes where our eyeball-norm method suggests there is an anomalous data point as well as in some places where our eyeballs are not accurate.

Below we print and plot any data points with scores greater than 3 standard deviations (approx 99.9th percentile) from the mean score.

```
[5]: score_mean = taxi_data['score'].mean()
score_std = taxi_data['score'].std()
score_cutoff = score_mean + 3 * score_std
anomalies = taxi_data_subset[taxi_data_subset['score'] > score_cutoff]
anomalies
```

The following is a list of known anomalous events which occurred in New York City within this timeframe:

```
[6]:
```





Amazon SageMaker Pipelines

1

Automatización de soluciones end to end

2

Trazabilidad

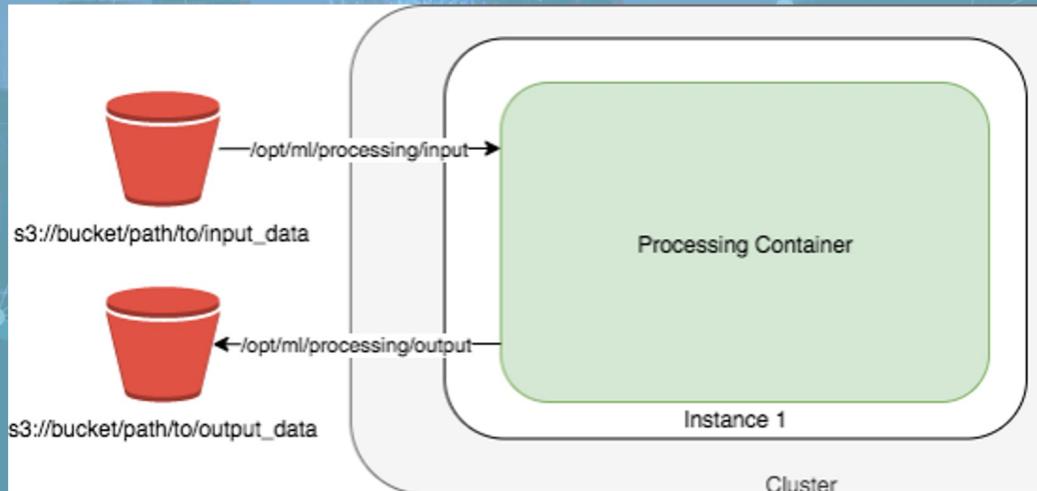
3

CI/CD

4

Reutilizables y escalables

Trabajos de Amazon SageMaker



1. Copia entradas desde s3 al contenedor (datos y código)
2. Ejecuta el código
3. Copia el resultado del contenedor a s3

Taller



<https://bit.ly/3ivvwbe>

Código hash:
5578-1f3eba79f4-f3



¡Gracias!

Sebastián Sandoval García

