Microsoft Sentinel

Make your life easier: **use machine learning!**

wortell

@jeroenniesen

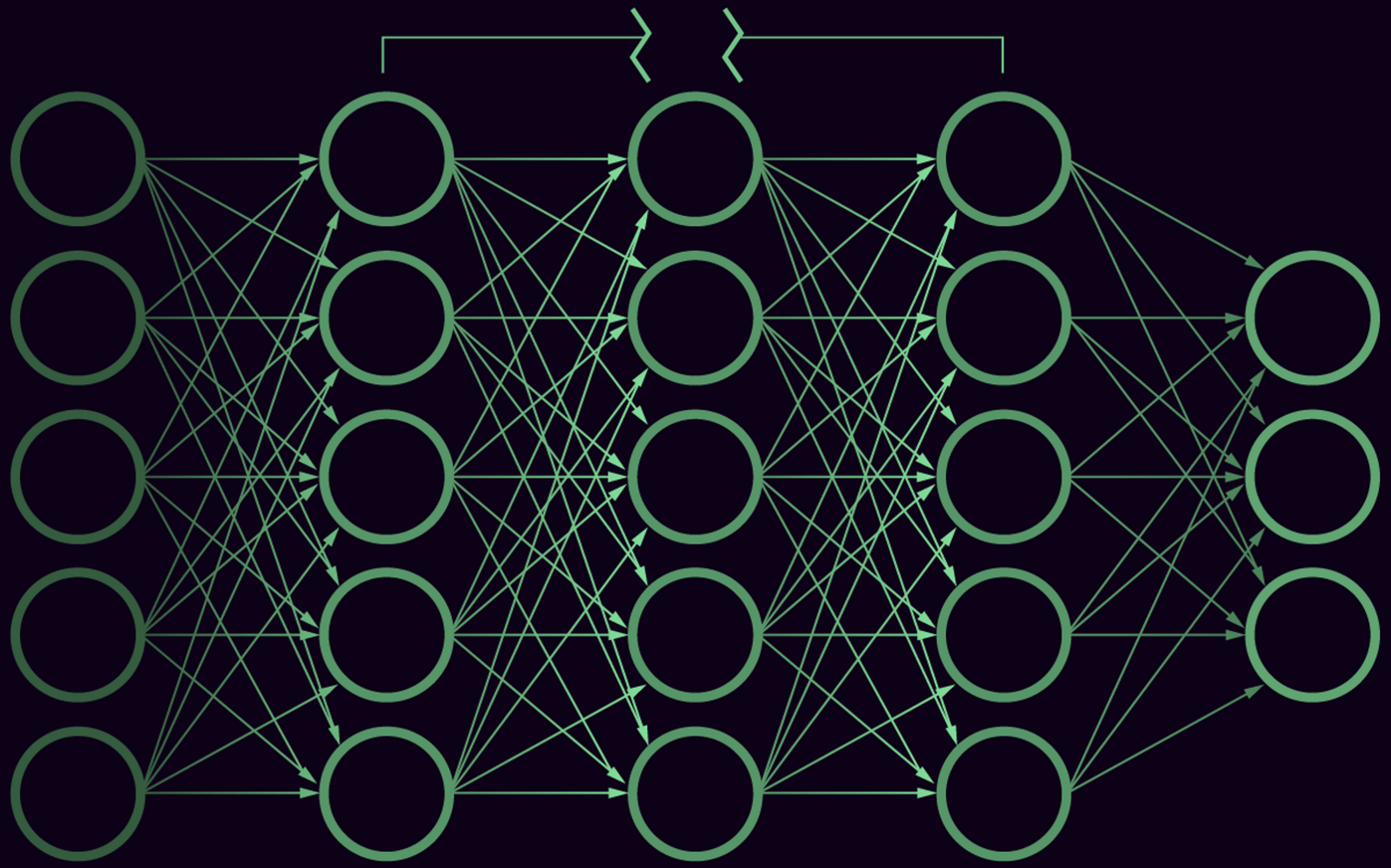
wortell

wortell

Input Layer

Hidden layers

Output Layer

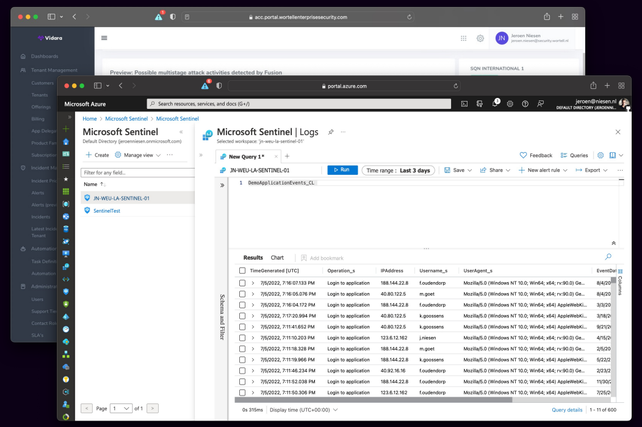


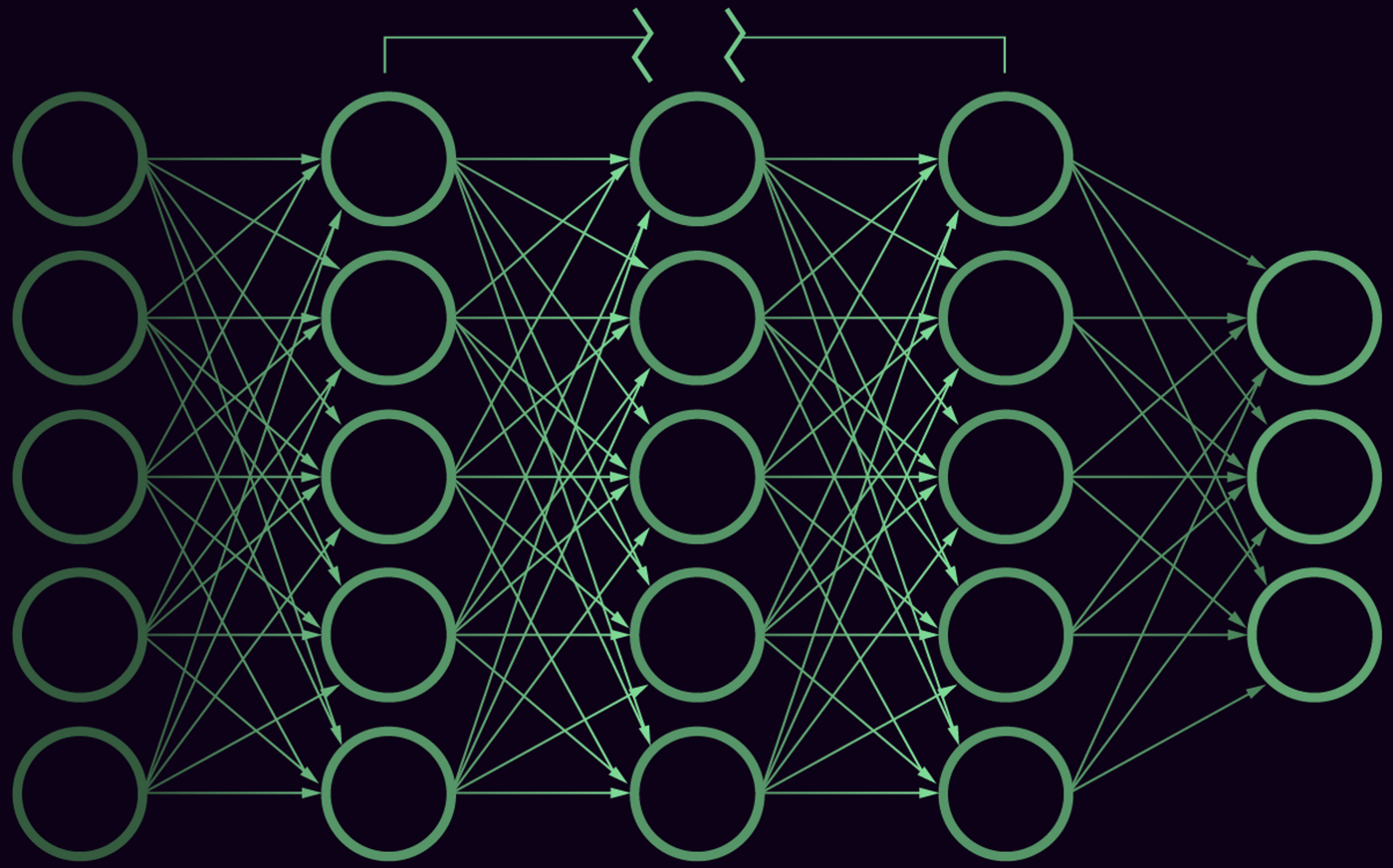
wortell

Input Layer

Hidden layers

Output Layer





Isolate

Reset password

Delete scheduled task

wortell

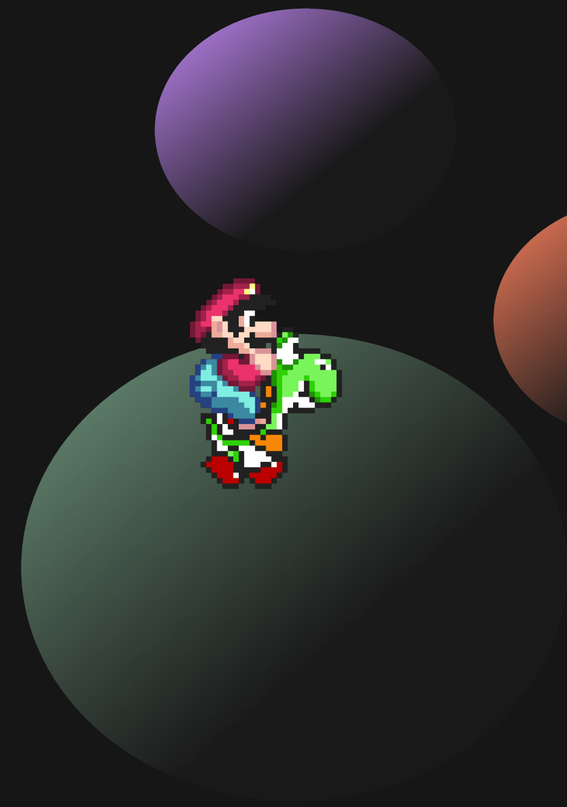
Artificial Intelligence and Machine Learning

- **Artificial Intelligence (AI)**

  Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by animals including humans.

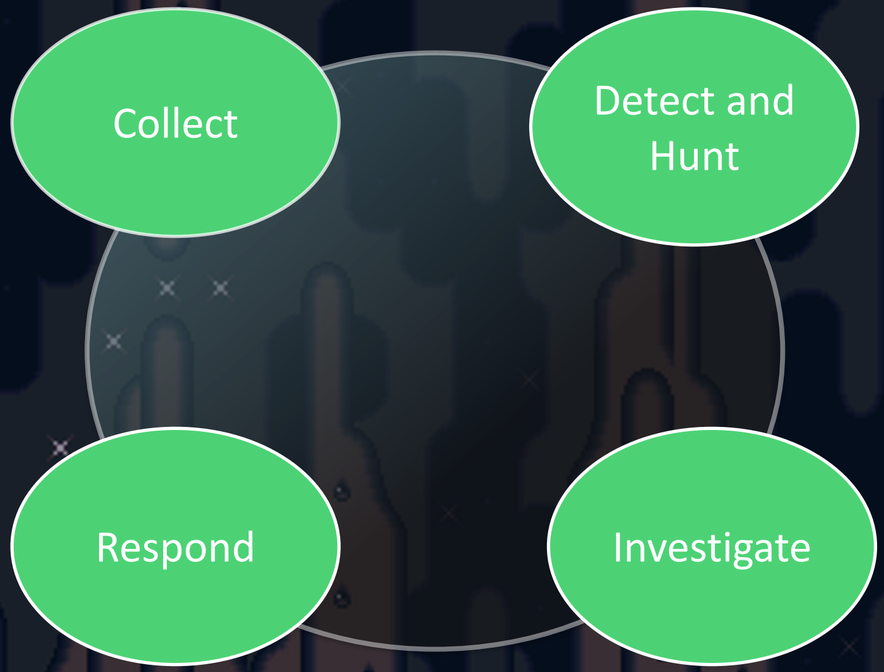
- **Machine Learning (ML)**

  Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. Machine learning is a subset of artificial intelligence.



wortell

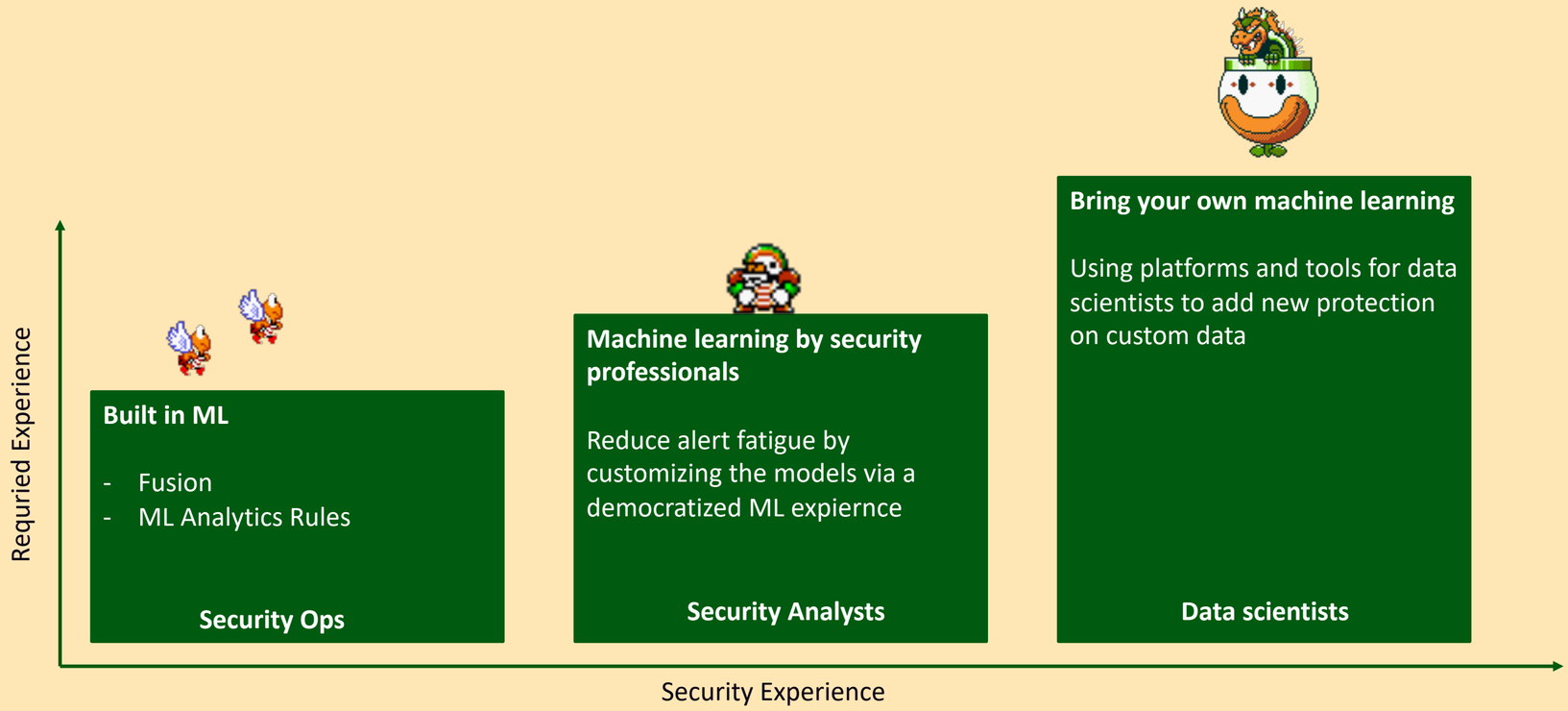
Azure Sentinel Key Capabilities



Machine learning at its best

wortell

Using machine learning in your security operations center

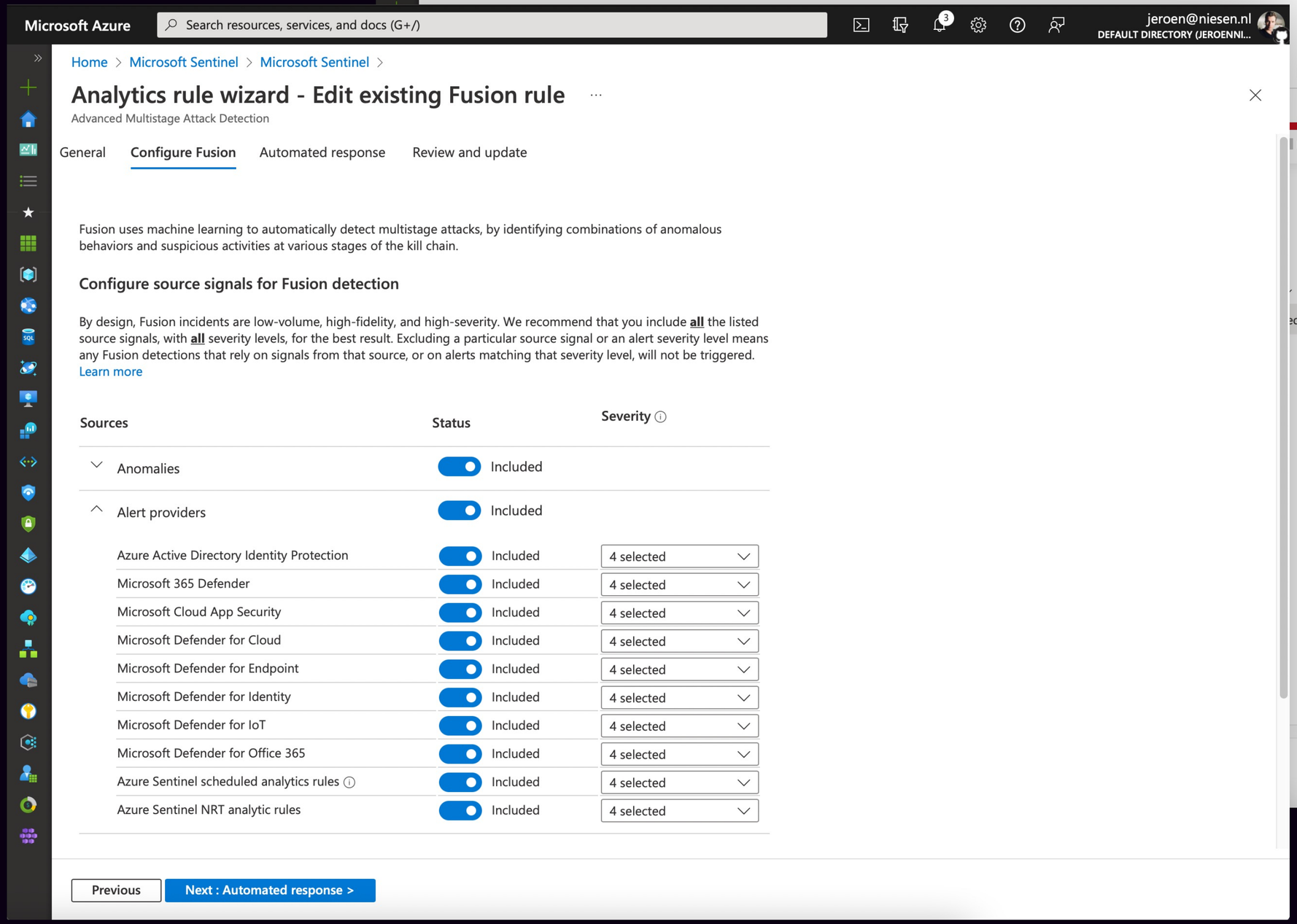


wortell

Fusion – Advanced Multistage Attack Detection

- Combining multiple activities into high fidelity incidents (red)
- Evaluate signals from multiple products to produce actionalble incidents
- Improve signal-to-noise while maintaining a low false positive rate
- Includes 122 scenario’s at this moment

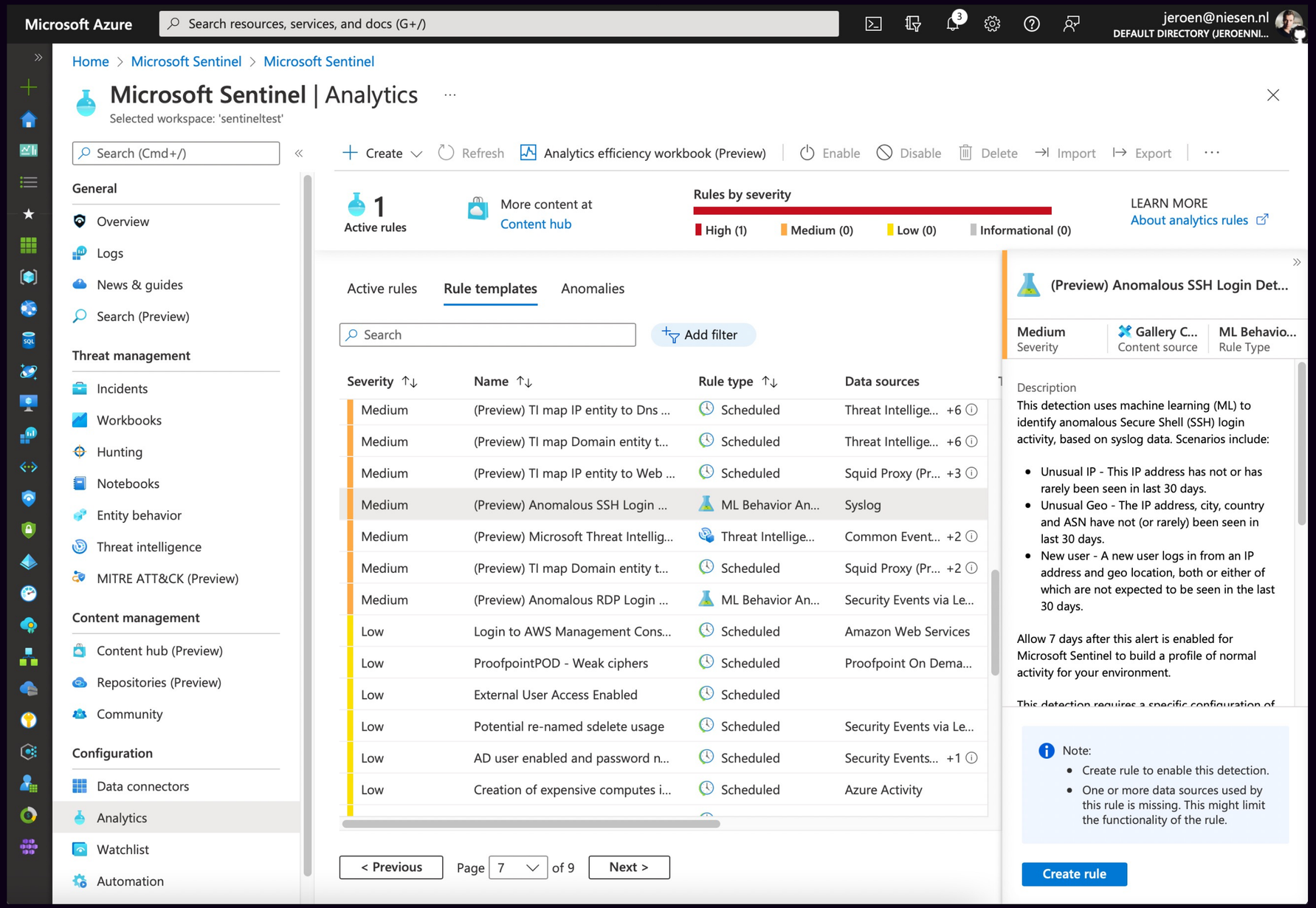
wortell


Microsoft Azure
Search resources, services, and docs (G+/)
jeroen@niesen.nl
DEFAULT DIRECTORY (JEROENNI...
Home > Microsoft Sentinel > Microsoft Sentinel
Analytics rule wizard - Edit existing Fusion rule
Advanced Multistage Attack Detection
General
Configure Fusion
Automated response
Review and update
Fusion uses machine learning to automatically detect multistage attacks, by identifying combinations of anomalous behaviors and suspicious activities at various stages of the kill chain.
Configure source signals for Fusion detection
By design, Fusion incidents are low-volume, high-fidelity, and high-severity. We recommend that you include all the listed source signals, with all severity levels, for the best result. Excluding a particular source signal or an alert severity level means any Fusion detections that rely on signals from that source, or on alerts matching that severity level, will not be triggered.
Learn more
Sources
Status
Severity
Anomalies
Included
Alert providers
Included
Azure Active Directory Identity Protection
Microsoft 365 Defender
Microsoft Cloud App Security
Microsoft Defender for Cloud
Microsoft Defender for Endpoint
Microsoft Defender for Identity
Microsoft Defender for IoT
Microsoft Defender for Office 365
Azure Sentinel scheduled analytics rules
Azure Sentinel NRT analytic rules
4 selected
Previous
Next : Automated response >
Delete
Import
Export
LEARN MORE
About analytics rules
Informational (0)
Advanced Multistage Attack Detecti...
High
Severity
Custom
Content source
Enabled
Status
Id
BuiltInFusion
Description
Microsoft Sentinel uses Fusion, a correlation engine based on scalable machine learning algorithms, to automatically detect multistage attacks by identifying combinations of anomalous behaviors and suspicious activities that are observed at various stages of the kill chain. On the basis of these discoveries, Azure Sentinel generates incidents that would otherwise be very difficult to catch. By design, these incidents are low-volume, high-fidelity, and high-severity, which is why this detection is turned ON by default.
Since Fusion correlates multiple signals from various products to detect advanced multistage attacks, successful Fusion detections are presented as Fusion incidents on the Microsoft Sentinel Incidents page. This rule covers the following detections:
Fusion for emerging threats
Fusion for ransomware
Scenario-based Fusion detections (122 scenarios)
Edit

Built-in SSH/RDP Login ML analytics rules

- Build user profiles based on two weeks
- Runs on streaming mode; time to detection is a few seconds

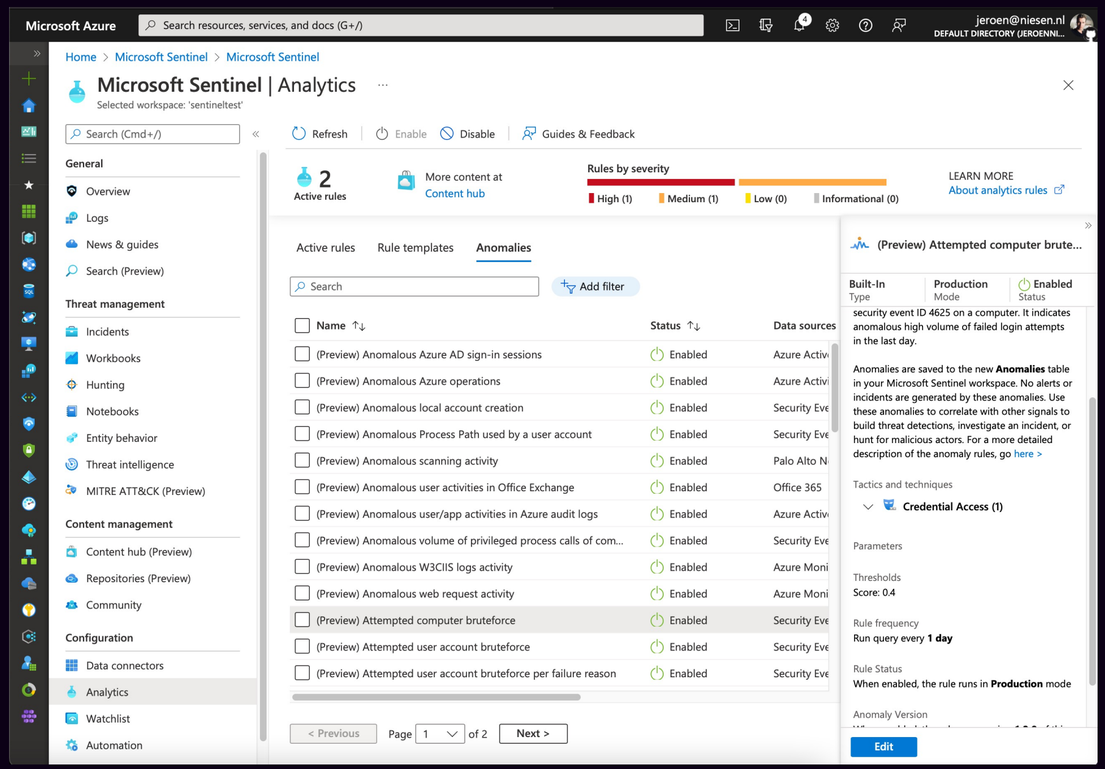
wortell


Microsoft Azure
Search resources, services, and docs (G+/)
jeroen@niesen.nl
DEFAULT DIRECTORY (JEROENNI...
Home > Microsoft Sentinel > Microsoft Sentinel
Microsoft Sentinel | Analytics
Selected workspace: 'sentineltest'
Search (Cmd+/)
General
Overview
Logs
News & guides
Search (Preview)
Threat management
Incidents
Workbooks
Hunting
Notebooks
Entity behavior
Threat intelligence
MITRE ATT&CK (Preview)
Content management
Content hub (Preview)
Repositories (Preview)
Community
Configuration
Data connectors
Analytics
Watchlist
Automation
Create
Refresh
Analytics efficiency workbook (Preview)
Enable
Disable
Delete
Import
Export
1
Active rules
More content at
Content hub
Rules by severity
High (1)
Medium (0)
Low (0)
Informational (0)
LEARN MORE
About analytics rules
Active rules
Rule templates
Anomalies
Search
Add filter
Severity
Name
Rule type
Data sources
Medium (Preview) TI map IP entity to Dns ... Scheduled Threat Intellige... +6
Medium (Preview) TI map Domain entity t... Scheduled Threat Intellige... +6
Medium (Preview) TI map IP entity to Web ... Scheduled Squid Proxy (Pr... +3
Medium (Preview) Anomalous SSH Login ... ML Behavior An... Syslog
Medium (Preview) Microsoft Threat Intellig... Threat Intellige... Common Event... +2
Medium (Preview) TI map Domain entity t... Scheduled Squid Proxy (Pr... +2
Medium (Preview) Anomalous RDP Login ... ML Behavior An... Security Events via Le...
Low Login to AWS Management Cons... Scheduled Amazon Web Services
Low ProofpointPOD - Weak ciphers Scheduled Proofpoint On Dema...
Low External User Access Enabled Scheduled
Low Potential re-named sdelete usage Scheduled Security Events via Le...
Low AD user enabled and password n... Scheduled Security Events... +1
Low Creation of expensive computes i... Scheduled Azure Activity
< Previous
Page 7 of 9
Next >
(Preview) Anomalous SSH Login Det...
Medium
Severity
Gallery C...
Content source
ML Behavio...
Rule Type
Description
This detection uses machine learning (ML) to identify anomalous Secure Shell (SSH) login activity, based on syslog data. Scenarios include:
Unusual IP - This IP address has not or has rarely been seen in last 30 days.
Unusual Geo - The IP address, city, country and ASN have not (or rarely) been seen in last 30 days.
New user - A new user logs in from an IP address and geo location, both or either of which are not expected to be seen in the last 30 days.
Allow 7 days after this alert is enabled for Microsoft Sentinel to build a profile of normal activity for your environment.
Note:
Create rule to enable this detection.
One or more data sources used by this rule is missing. This might limit the functionality of the rule.
Create rule

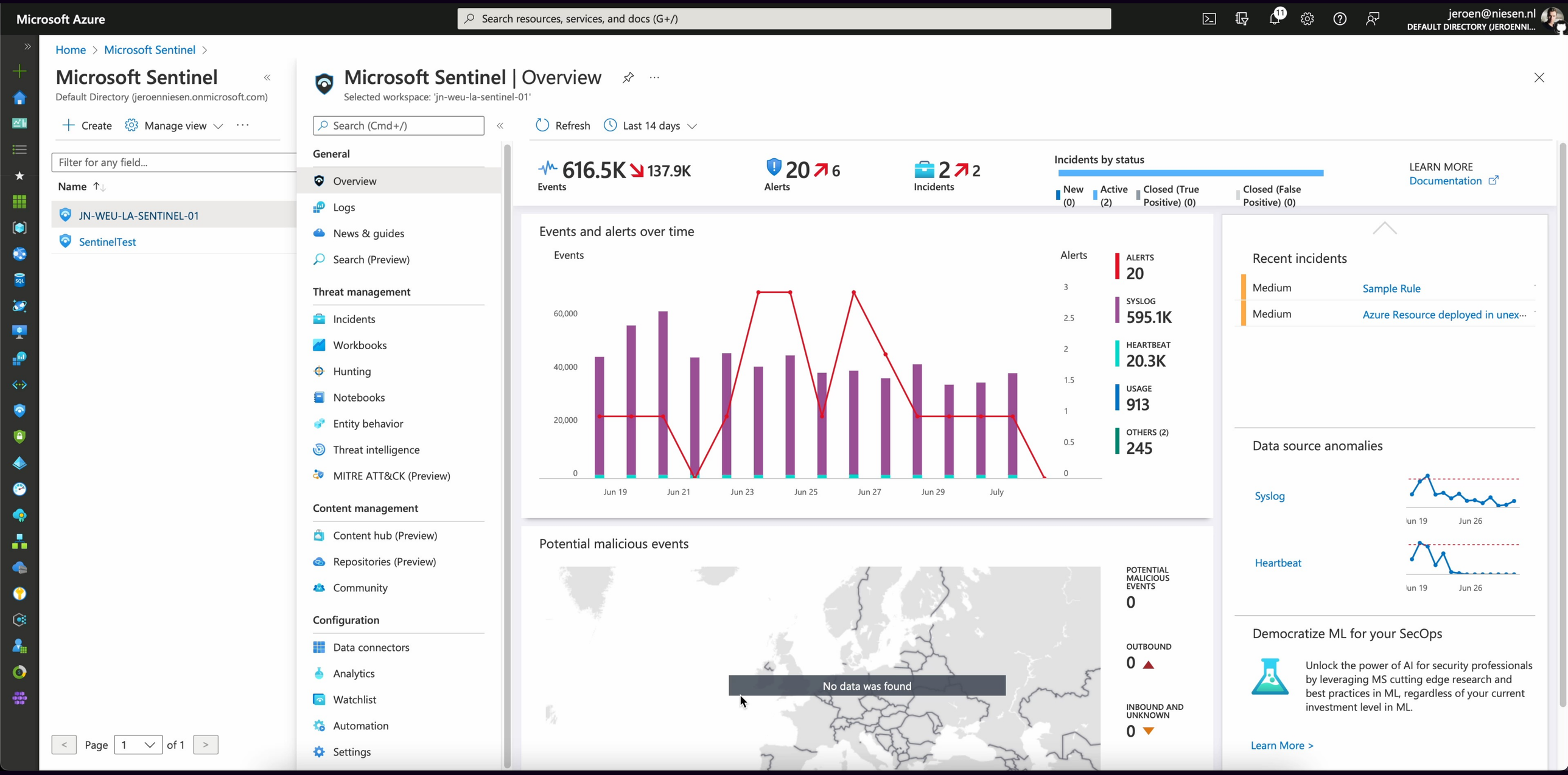
wortell

Anomaly Rules/Detections

- **Machine learning algorithms** that are being used to detect a wide variaty of anomalies.
- Anomalies are saved in the **Anomalies** table
- Algorithms can be **influenced with parameters** (e.g. threshold, scoring, etc.)



wortell


Microsoft Azure
Search resources, services, and docs (G+/)
jeroen@niesen.nl
DEFAULT DIRECTORY (JEROENNI...
Home > Microsoft Sentinel >
Microsoft Sentinel
Default Directory (jeroenniesen.onmicrosoft.com)
Create
Manage view
Filter for any field...
Name
JN-WEU-LA-SENTINEL-01
SentinelTest
Page 1 of 1
Microsoft Sentinel | Overview
Selected workspace: 'jn-weu-la-sentinel-01'
Search (Cmd+/)
Refresh
Last 14 days
General
Overview
Logs
News & guides
Search (Preview)
Threat management
Incidents
Workbooks
Hunting
Notebooks
Entity behavior
Threat intelligence
MITRE ATT&CK (Preview)
Content management
Content hub (Preview)
Repositories (Preview)
Community
Configuration
Data connectors
Analytics
Watchlist
Automation
Settings
616.5K 137.9K
Events
20 6
Alerts
2 2
Incidents
Incidents by status
New (0)
Active (2)
Closed (True Positive) (0)
Closed (False Positive) (0)
LEARN MORE
Documentation
Events and alerts over time
Events
Alerts
60,000
40,000
20,000
0
3
2.5
2
1.5
1
0.5
0
Jun 19
Jun 21
Jun 23
Jun 25
Jun 27
Jun 29
July
ALERTS
20
SYSLOG
595.1K
HEARTBEAT
20.3K
USAGE
913
OTHERS (2)
245
Potential malicious events
No data was found
POTENTIAL MALICIOUS EVENTS
0
OUTBOUND
0
INBOUND AND UNKNOWN
0
Recent incidents
Medium
Sample Rule
Medium
Azure Resource deployed in unex...
Data source anomalies
Syslog
Heartbeat
Jun 19
Jun 26
Democratize ML for your SecOps
Unlock the power of AI for security professionals by leveraging MS cutting edge research and best practices in ML, regardless of your current investment level in ML.
Learn More >

Bring your own ML

Why would you create your own ML models?

- When using custom data
- To detect anomalies that you would not detect using a regular KQL query

wortell

The Adventure

- A company has 4 employees that use an application. This application has a custom user module and logs all the sign-ins. The employees always work on a Windows device from a limited set of (public IP) addresses.
- An adversary tries to attack the application that is used by the employees. The adversary is already in the posession of the user credentials.

Are we able to determine when the adversary is signing in?

wortell

Path to success

wortell

Steps to create a model

1. Understand data
2. Have access to the data
3. Transform data so it can be used in machine learning
4. Split the dataset into a training dataset and validation dataset
5. Select your machine learning model
6. Train your ML model
7. Validate your ML model
8. Use your model

wortell

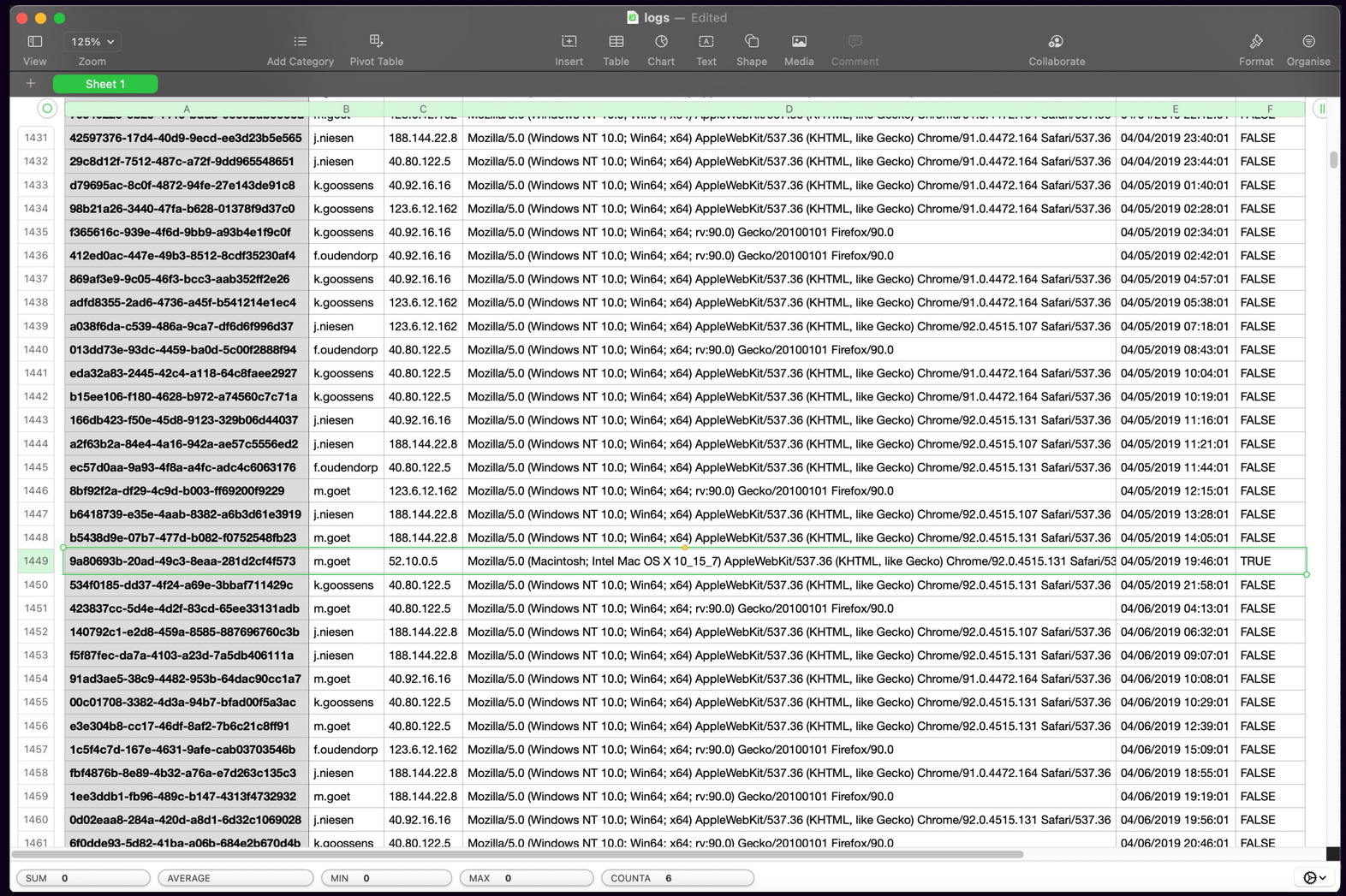
Understanding data

- What data is available?
- How much data is available?
- Do you have access to the values that you try to predict?
- What format will the data be in?
- Where does it reside?
- Which fields can influence the value dat you try to predict?

wortell

The dataset

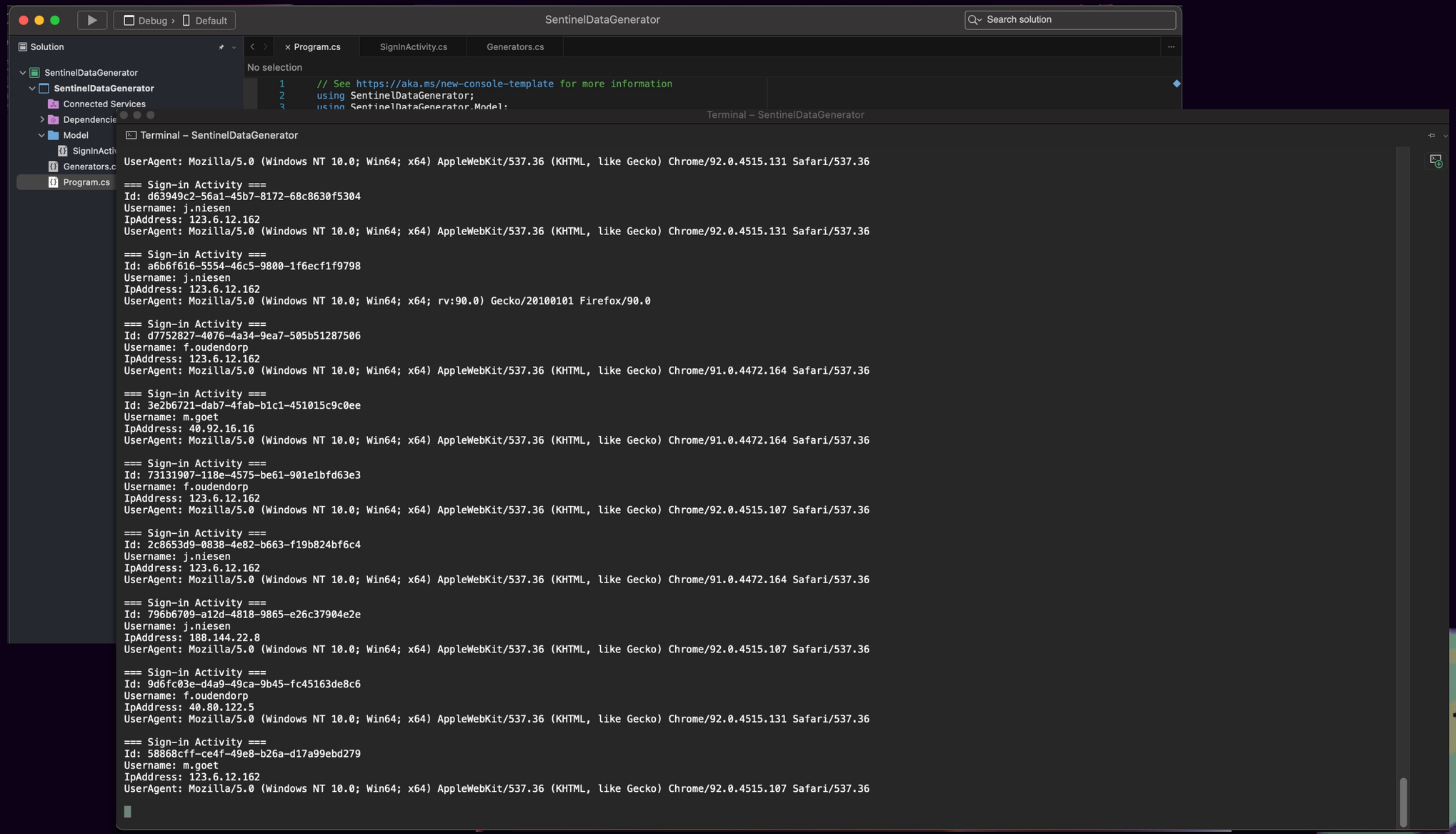
- **Needs to be labeled/classified**. In order to develop, train and test your model, you need to have a classified/labeled set of data.



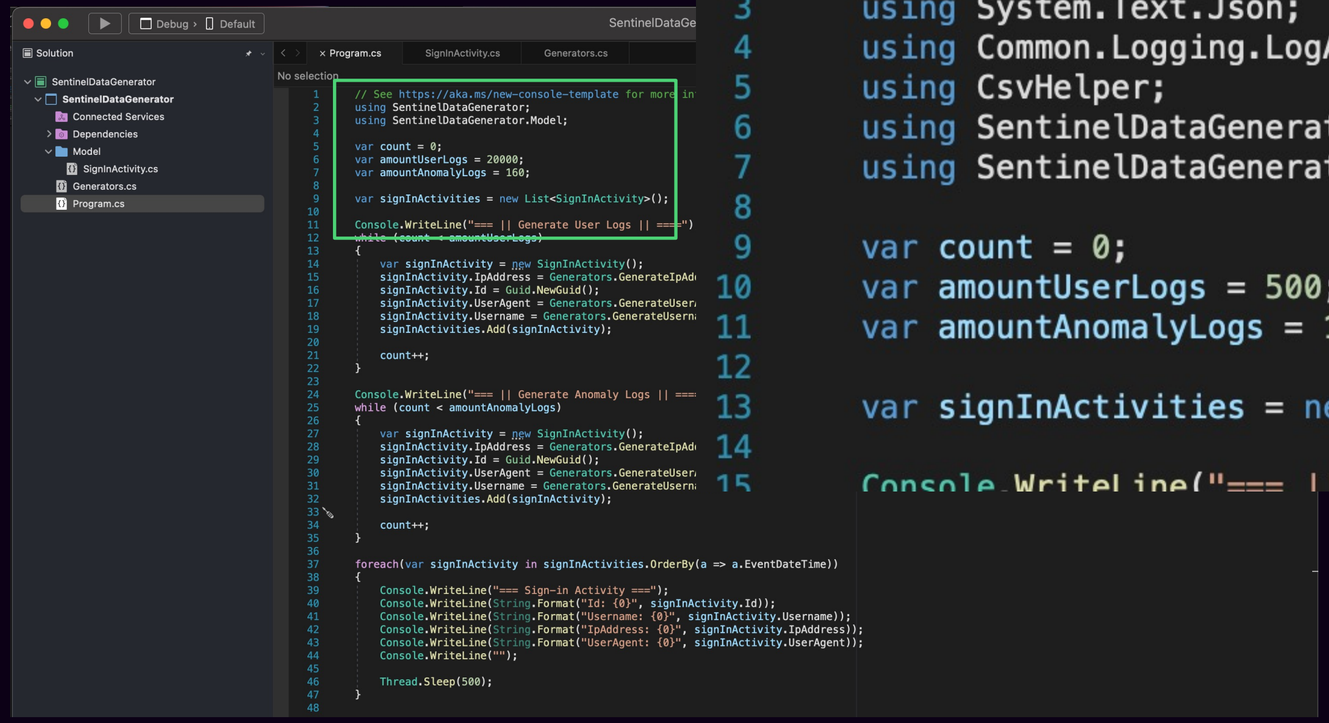
	A	B	C	D	E	F
1431	42597376-17d4-40d9-9ecd-ee3d23b5e565	j.niesen	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/04/2019 23:40:01	FALSE
1432	29c8d12f-7512-487c-a72f-9dd965548651	j.niesen	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/04/2019 23:44:01	FALSE
1433	d79695ac-8c0f-4872-94fe-27e143de91c8	k.goossens	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/05/2019 01:40:01	FALSE
1434	98b21a26-3440-47fa-b628-01378f9d37c0	k.goossens	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/05/2019 02:28:01	FALSE
1435	f365616c-939e-4f6d-9bb9-a93b4e1f9c0f	k.goossens	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/05/2019 02:34:01	FALSE
1436	412ed0ac-447e-49b3-8512-8cdf35230af4	f.oudendorp	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/05/2019 02:42:01	FALSE
1437	869af3e9-9c05-46f3-bcc3-aab352ff2e26	k.goossens	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/05/2019 04:57:01	FALSE
1438	adfd8355-2ad6-4736-a45f-b541214e1ec4	k.goossens	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/05/2019 05:38:01	FALSE
1439	a038f6da-c539-486a-9ca7-df6d6f996d37	j.niesen	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	04/05/2019 07:18:01	FALSE
1440	013dd73e-93dc-4459-ba0d-5c00f2888f94	f.oudendorp	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/05/2019 08:43:01	FALSE
1441	eda32a83-2445-42c4-a118-64c8faee2927	k.goossens	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/05/2019 10:04:01	FALSE
1442	b15ee106-f180-4628-b972-a74560c7c71a	k.goossens	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/05/2019 10:19:01	FALSE
1443	166db423-f50e-45d8-9123-329b06d44037	j.niesen	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/05/2019 11:16:01	FALSE
1444	a2f63b2a-84e4-4a16-942a-ae57c5556ed2	j.niesen	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	04/05/2019 11:21:01	FALSE
1445	ec57d0aa-9a93-4f8a-a4fc-adc4c6063176	f.oudendorp	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/05/2019 11:44:01	FALSE
1446	8bf92f2a-df29-4c9d-b003-ff69200f9229	m.goet	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/05/2019 12:15:01	FALSE
1447	b6418739-e35e-4aab-8382-a6b3d61e3919	j.niesen	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	04/05/2019 13:28:01	FALSE
1448	b5438d9e-07b7-477d-b082-f0752548fb23	m.goet	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/05/2019 14:05:01	FALSE
1449	9a80693b-20ad-49c3-8eaa-281d2cf4f573	m.goet	52.10.0.5	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/53	04/05/2019 19:46:01	TRUE
1450	534f0185-dd37-4f24-a69e-3bbaf711429c	k.goossens	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/05/2019 21:58:01	FALSE
1451	423837cc-5d4e-4d2f-83cd-65ee33131adb	m.goet	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/06/2019 04:13:01	FALSE
1452	140792c1-e2d8-459a-8585-887696760c3b	j.niesen	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	04/06/2019 06:32:01	FALSE
1453	f5f87fec-da7a-4103-a23d-7a5db406111a	j.niesen	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/06/2019 09:07:01	FALSE
1454	91ad3ae5-38c9-4482-953b-64dac90cc1a7	m.goet	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/06/2019 10:08:01	FALSE
1455	00c01708-3382-4d3a-94b7-bfad00f5a3ac	k.goossens	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/06/2019 10:29:01	FALSE
1456	e3e304b8-cc17-46df-8af2-7b6c21c8ff91	m.goet	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/06/2019 12:39:01	FALSE
1457	1c5f4c7d-167e-4631-9afe-cab03703546b	f.oudendorp	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/06/2019 15:09:01	FALSE
1458	fbf4876b-8e89-4b32-a76a-e7d263c135c3	j.niesen	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	04/06/2019 18:55:01	FALSE
1459	1ee3ddb1-fb96-489c-b147-4313f4732932	m.goet	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/06/2019 19:19:01	FALSE
1460	0d02eaa8-284a-420d-a8d1-6d32c1069028	j.niesen	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/06/2019 19:56:01	FALSE
1461	6f0dde93-5d82-41ba-a06b-684e2b670d4b	k.goossens	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/06/2019 20:46:01	FALSE

wortell

Sentinel data generator



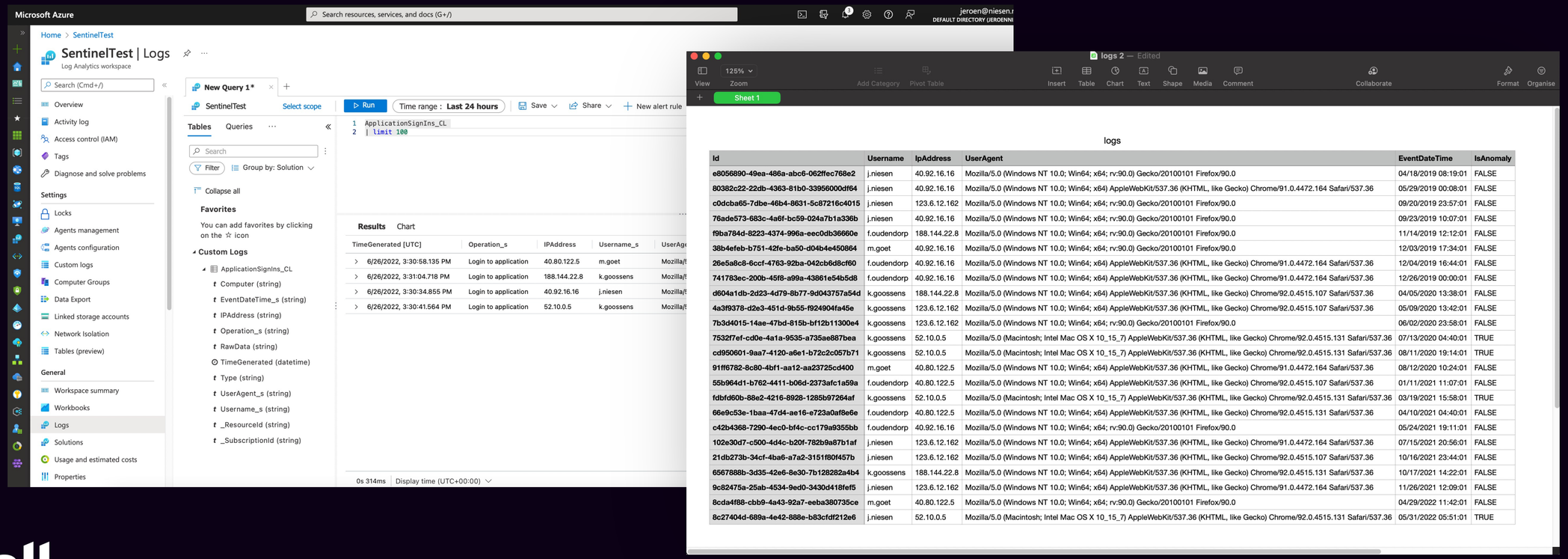
Sentinel data generator




```
// See https://aka.ms/new-console-template for more information
using System.Globalization;
using System.Text.Json;
using Common.Logging.LogAnalytics;
using CsvHelper;
using SentinelDataGenerator;
using SentinelDataGenerator.Model;

var count = 0;
var amountUserLogs = 500;
var amountAnomalyLogs = 100;

var signInActivities = new List<SignInActivity>();

Console.WriteLine("=== || Generate User Logs || ====");
```

Sentinel data generator

- **Output to Microsoft Sentinel & CSV**. The CSV file can help developing, training and validating your machine learning model.



Id	Username	IpAddress	UserAgent	EventDateTime	IsAnomaly
e8056890-49ea-486a-abc6-062ffec768e2	j.niesen	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/18/2019 08:19:01	FALSE
80382c22-22db-4363-81b0-33956000df64	j.niesen	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	05/29/2019 00:08:01	FALSE
c0dcba65-7dbe-46b4-8631-5c87216c4015	j.niesen	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	09/20/2019 23:57:01	FALSE
76ade573-683c-4a6f-bc59-024a7b1a336b	j.niesen	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	09/23/2019 10:07:01	FALSE
f9ba784d-8223-4374-996a-eec0db36660e	f.oudendorp	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	11/14/2019 12:12:01	FALSE
38b4efeb-b751-42fe-ba50-d04b4e450864	m.goet	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	12/03/2019 17:34:01	FALSE
26e5a8c8-6ccf-4763-92ba-042cb6d8cf60	f.oudendorp	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	12/04/2019 16:44:01	FALSE
741783ec-200b-45f8-a99a-43861e54b5d8	f.oudendorp	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	12/26/2019 00:00:01	FALSE
d604a1db-2d23-4d79-8b77-9d043757a54d	k.goossens	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	04/05/2020 13:38:01	FALSE
4a3f9378-d2e3-451d-9b55-f924904fa45e	k.goossens	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	05/09/2020 13:42:01	FALSE
7b3d4015-14ae-47bd-815b-bf12b11300e4	k.goossens	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	06/02/2020 23:58:01	FALSE
7532f7ef-cd0e-4a1a-9535-a735ae887bea	k.goossens	52.10.0.5	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	07/13/2020 04:40:01	TRUE
cd950601-9aa7-4120-a6e1-b72c2c057b71	k.goossens	52.10.0.5	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	08/11/2020 19:14:01	TRUE
91ff6782-8c80-4bf1-aa12-aa23725cd400	m.goet	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	08/12/2020 10:24:01	FALSE
55b964d1-b762-4411-b06d-2373afc1a59a	f.oudendorp	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	01/11/2021 11:07:01	FALSE
fdbfd60b-88e2-4216-8928-1285b97264af	k.goossens	52.10.0.5	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	03/19/2021 15:58:01	TRUE
66e9c53e-1baa-47d4-ae16-e723a0af8e6e	f.oudendorp	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	04/10/2021 04:40:01	FALSE
c42b4368-7290-4ec0-bf4c-cc179a9355bb	f.oudendorp	40.92.16.16	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	05/24/2021 19:11:01	FALSE
102e30d7-c500-4d4c-b20f-782b9a87b1af	j.niesen	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	07/15/2021 20:56:01	FALSE
21db273b-34cf-4ba6-a7a2-3151f80f457b	j.niesen	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.107 Safari/537.36	10/16/2021 23:44:01	FALSE
6567888b-3d35-42e6-8e30-7b128282a4b4	k.goossens	188.144.22.8	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	10/17/2021 14:22:01	FALSE
9c82475a-25ab-4534-9ed0-3430d418fef5	j.niesen	123.6.12.162	Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.164 Safari/537.36	11/26/2021 12:09:01	FALSE
8cda4f88-cbb9-4a43-92a7-eeba380735ce	m.goet	40.80.122.5	Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:90.0) Gecko/20100101 Firefox/90.0	04/29/2022 11:42:01	FALSE
8c27404d-689a-4e42-888e-b83cfdf212e6	j.niesen	52.10.0.5	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36	05/31/2022 05:51:01	TRUE

wortell


MORTON
KOOPA JR.
ROY
KOOPA
BOWSER
LEMMY
KOOPA
WENDY
O. KOOPA
KOOPA
LUDWIG
VON KOOPA
LARRY
KOOPA

wortell

Split the dataset (sampling)

- **Random samping**. This data sampling method protects the data modeling proces from bias to ddifferten possible characteristics; it however my have issues regarding to uneven distribution of data.
- **Stratified random sampling**. This sampling method selects data samples at random with specific parameters. It ensures data is distributed correctly.
- **Nonrandom sampling**. Usefull when you would like to use the most recent data for example

wortell

Selecting the model

wortell

Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution.
Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

Microsoft Azure

What do you want to do?

Extract information from text

Text Analytics

Derives high-quality information from text

Answers questions like: What info is in this text?

- Latent Dirichlet Allocation ← Unsupervised topic modeling, group texts that are similar
- Extract N-Gram Features from Text ← Creates a dictionary of n-grams from a column of free text
- Feature Hashing ← Converts text data to integer encoded features using the Vowpal Wabbit library
- Preprocess Text ← Performs cleaning operations on text, like removal of stop-words, case normalization
- Word2Vector ← Converts words to values for use in NLP tasks, like recommender, named entity recognition, machine translation

Predict between several categories

Multiclass Classification

Answers complex questions with multiple possible answers

Answers questions like: Is this A or B or C or D?

- Multiclass Logistic Regression ← Fast training times, linear model
- Multiclass Neural Network ← Accuracy, long training times
- Multiclass Decision Forest ← Accuracy, fast training times
- One-vs-All Multiclass ← Depends on the two-class classifier
- One-vs-One Multiclass ← Depends on binary classifier, less sensitive to an imbalanced dataset with larger complexity
- Multiclass Boosted Decision Tree ← Non-parametric, fast training times and scalable

Predict between two categories

Two-Class Classification

Answers simple two-choice questions, like yes or no, true or false

Answers questions like: Is this A or B?

- Two-Class Support Vector Machine ← Under 100 features, linear model
- Two-Class Averaged Perceptron ← Fast training, linear model
- Two-Class Decision Forest ← Accurate, fast training
- Two-Class Logistic Regression ← Fast training, linear model
- Two-Class Boosted Decision Tree ← Accurate, fast training, large memory footprint
- Two-Class Neural Network ← Accurate, long training times

Generate recommendations

Recommenders

Predicts what someone will be interested in

Answers the question: What will they be interested in?

- Use the Train Wide & Deep Recommender module ← Hybrid recommender, both collaborative filtering and content-based approach
- SVD Recommender ← Collaborative filtering, better performance with lower cost by reducing dimensionality

Predict values

Regression

Makes forecasts by estimating the relationship between values

Answers questions like: How much or how many?

- Fast Forest Quantile Regression ← Predicts a distribution
- Poisson Regression ← Predicts event counts
- Linear Regression ← Fast training, linear model
- Bayesian Linear Regression ← Linear model, small data sets
- Decision Forest Regression ← Accurate, fast training times
- Neural Network Regression ← Accurate, long training times
- Boosted Decision Tree Regression ← Accurate, fast training times, large memory footprint

Discover structure

Clustering

Separates similar data points into intuitive groups

Answers questions like: How is this organized?

- K-Means ← Unsupervised learning

Find unusual occurrences

Anomaly Detection

Identifies and predicts rare or unusual data points

Answers the question: Is this weird?

- One Class SVM ← Under 100 features, aggressive boundary
- PCA-Based Anomaly Detection ← Fast training times

Classify images

Image Classification

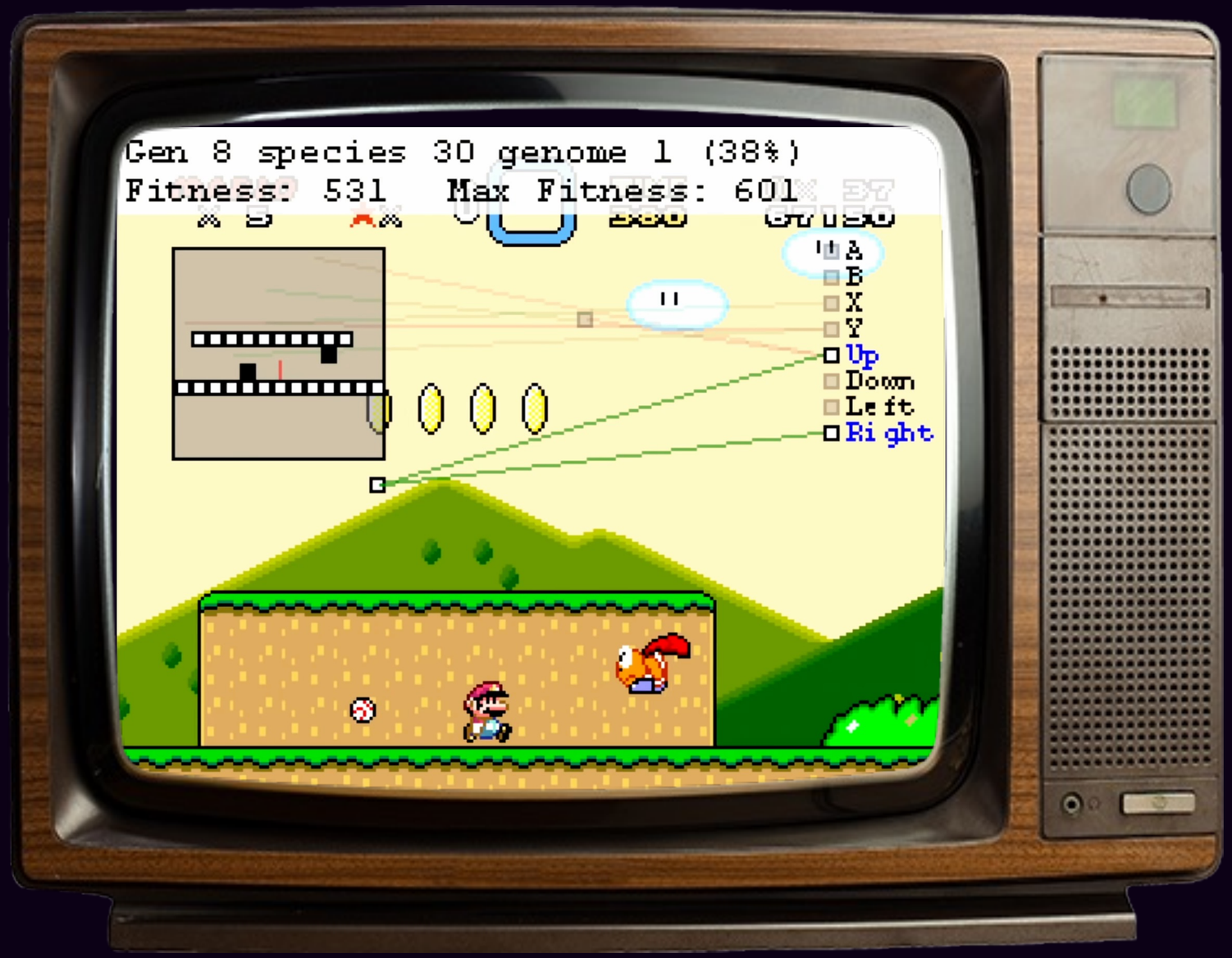
Classifies images with popular networks

Answers questions like: What does this image represent?

- ResNet ← Modern deep learning neural network
- DenseNet ← Modern deep learning neural network

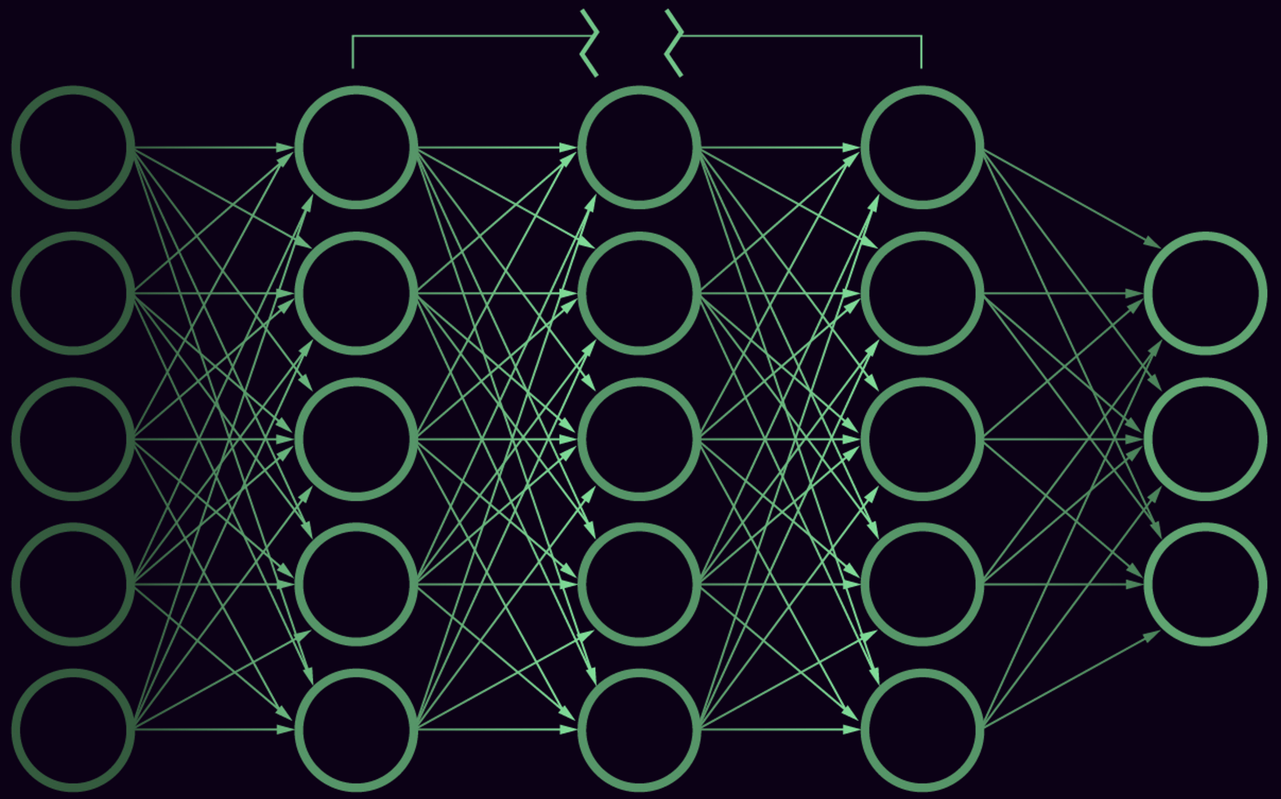
Nural Network Training

- Iterative train a model and learn from its mistakes
- Each iteration generates a new generation of the nural network which is smarter than is predecessor



wortell

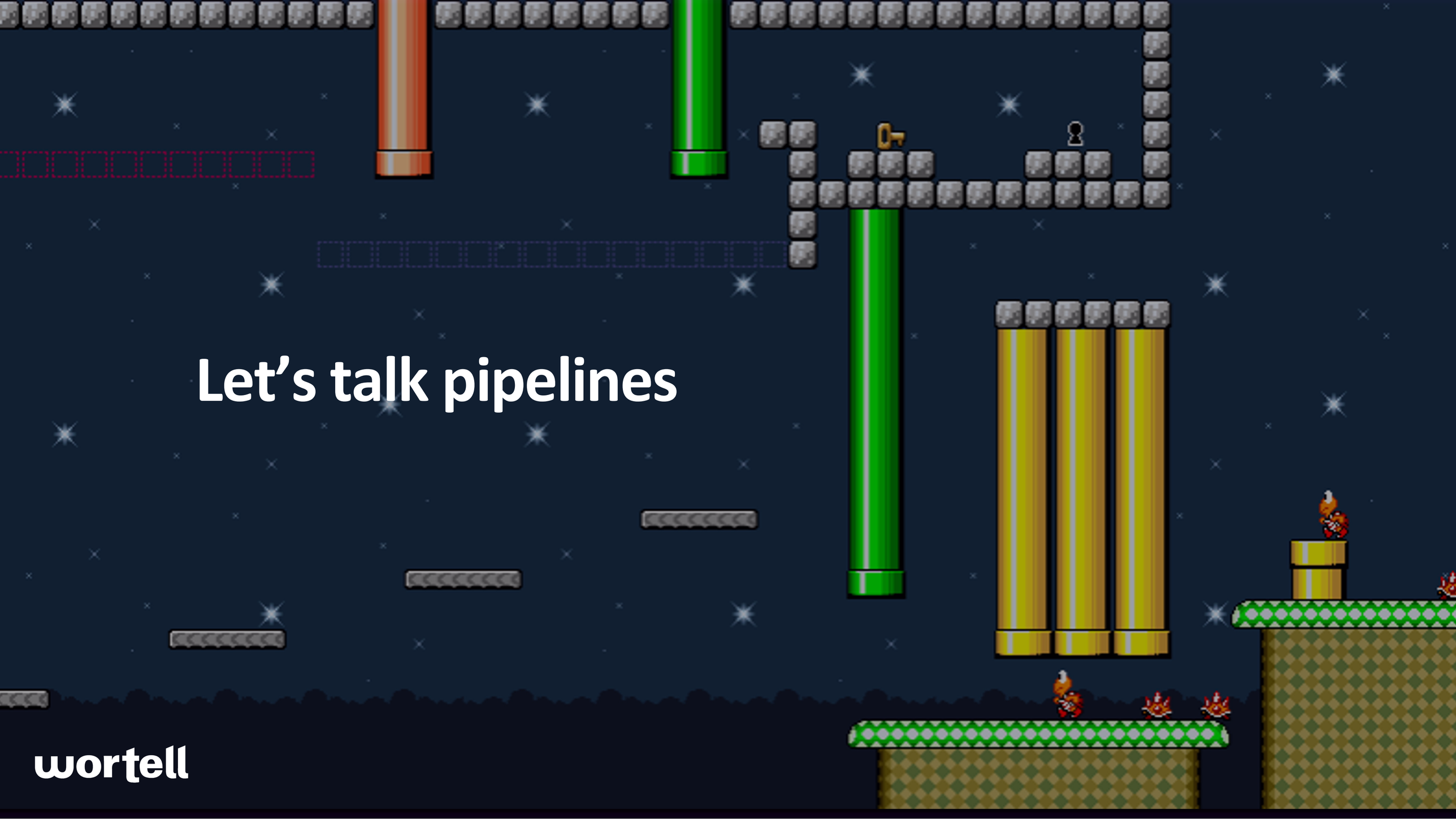
```
{
  “Username” : “j.niesen”
  “IpAddress”: “123.123.123.2”
 “UserAgent”: “Mozilla/5.0…”
}
```



TRUE (=adversary)

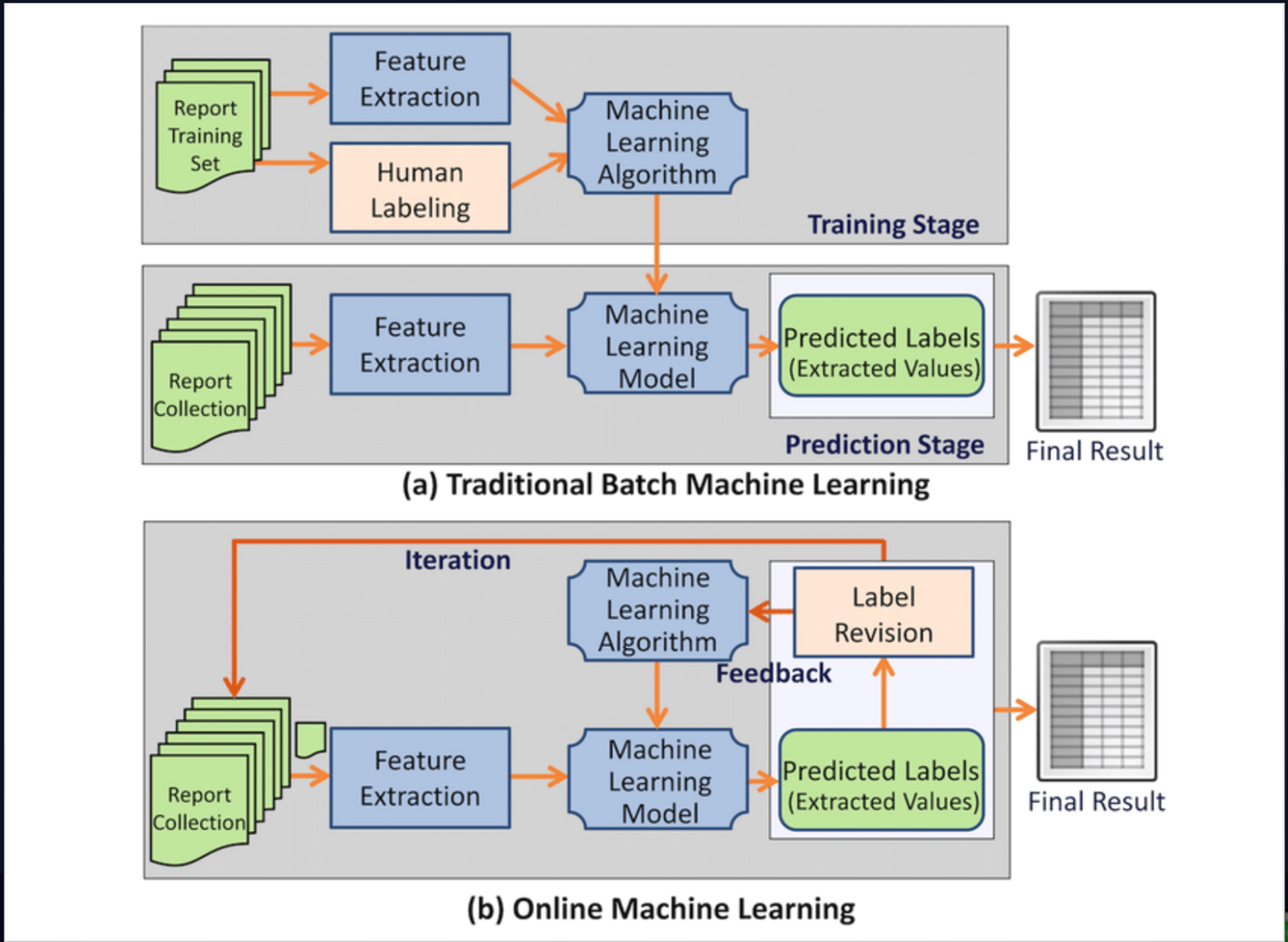
FALSE(=not a adversary)

wortell



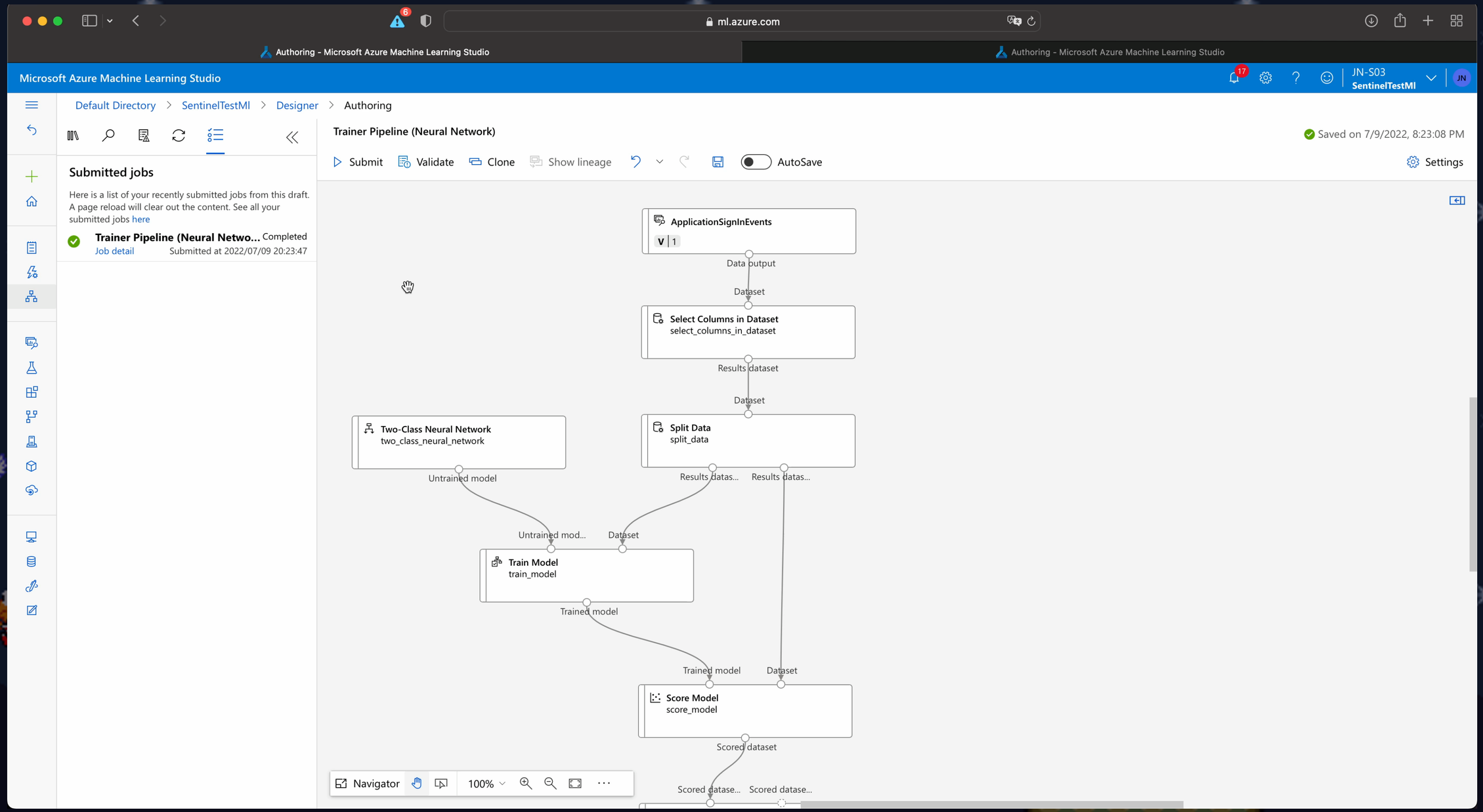
wortell

Architecture Decision: Online vs. Offline Pipeline



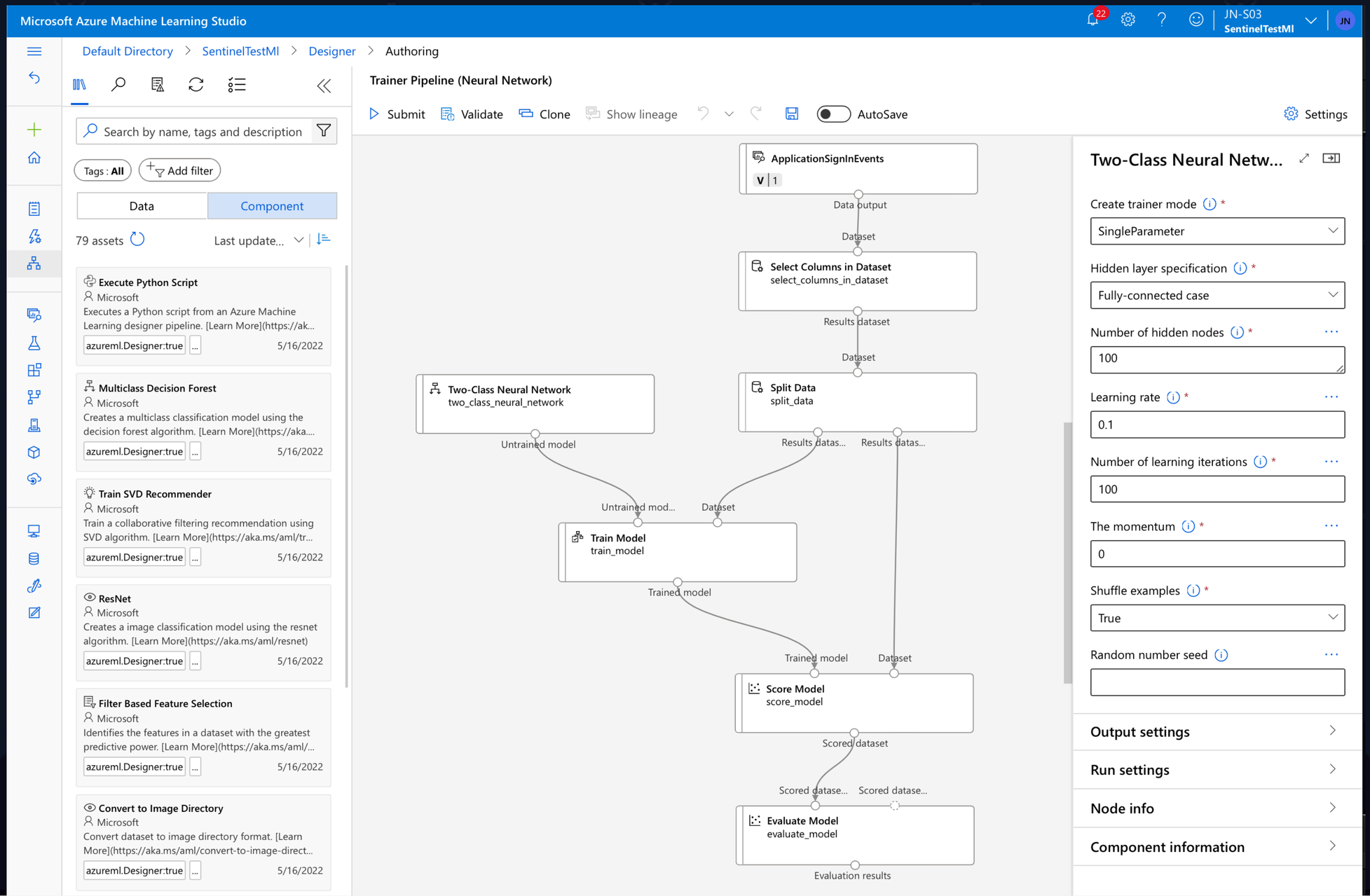
Train the model

wortell



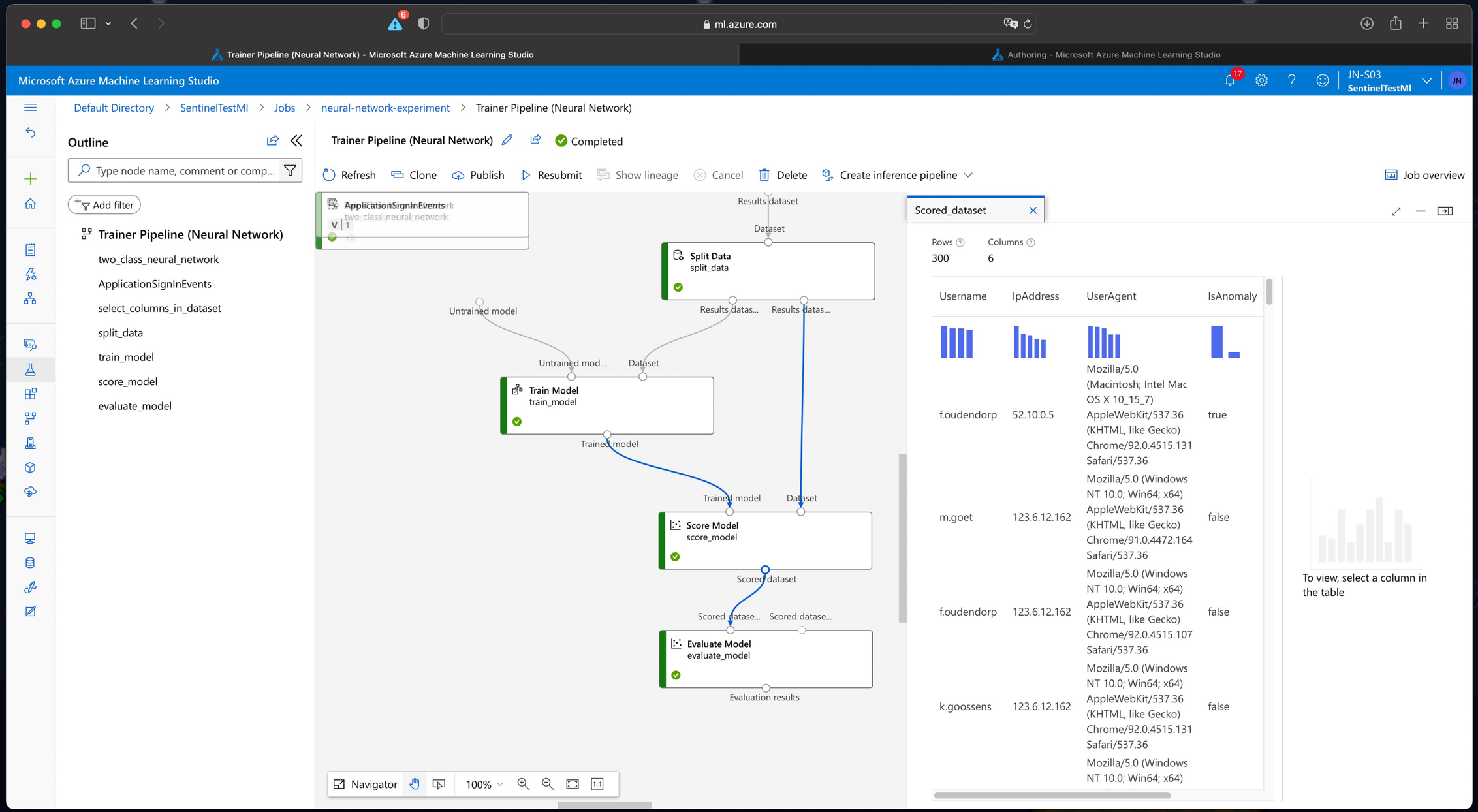
Train the model

wortell



Train the model

wortell



Pipelines in Azure Machine Learning

Types of pipelines

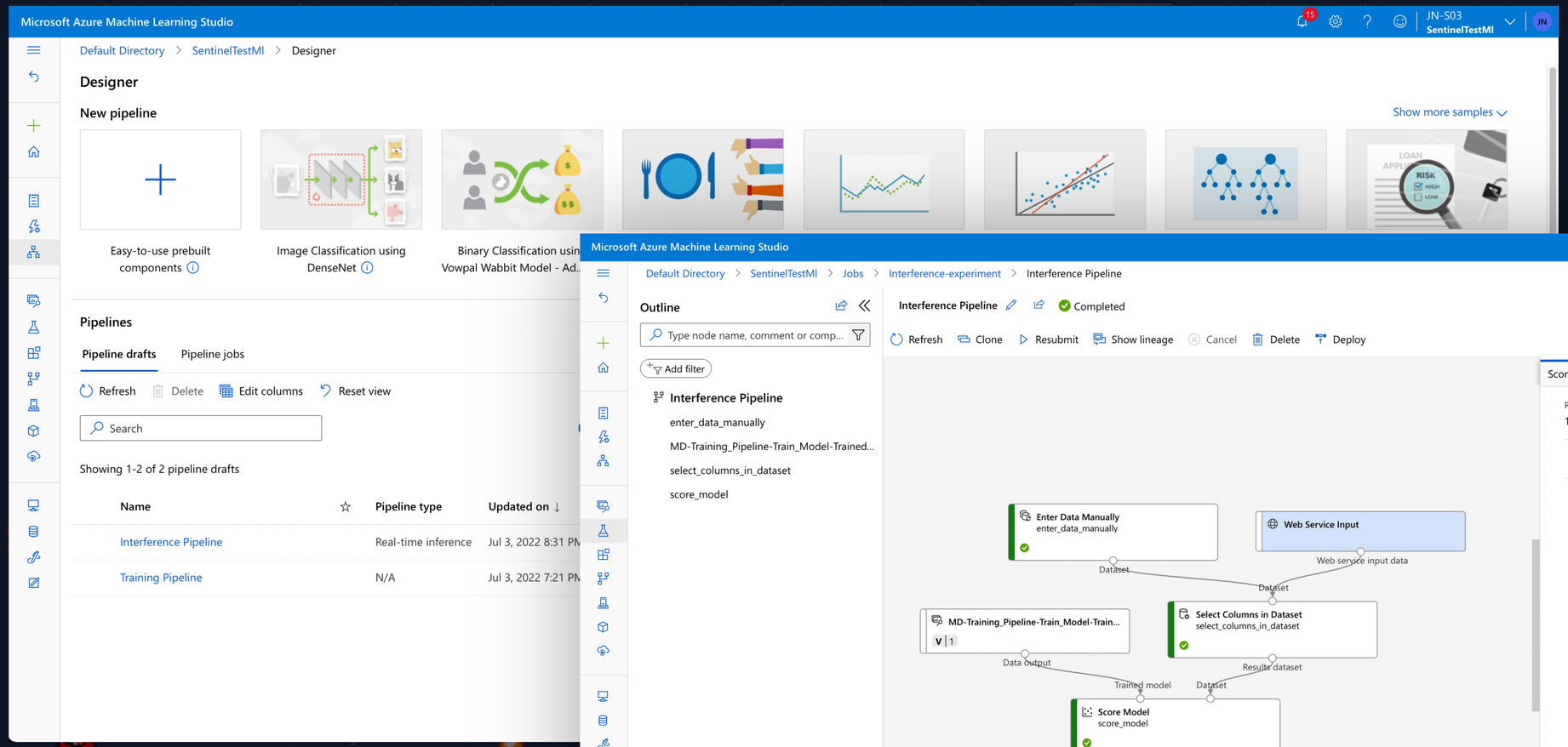
- **Regular**. Contains the flow to train the machine learning model
- **Interference Pipeline**. Contains the flow to form the template for a webservice to consume the trained model

Good to know

- **Pipelines run on compute.** Depending on the type of model, the amount of data, the compute can vary.
- **Pipelines use components and data** that are configured in your Azure ML workspace.

wortell

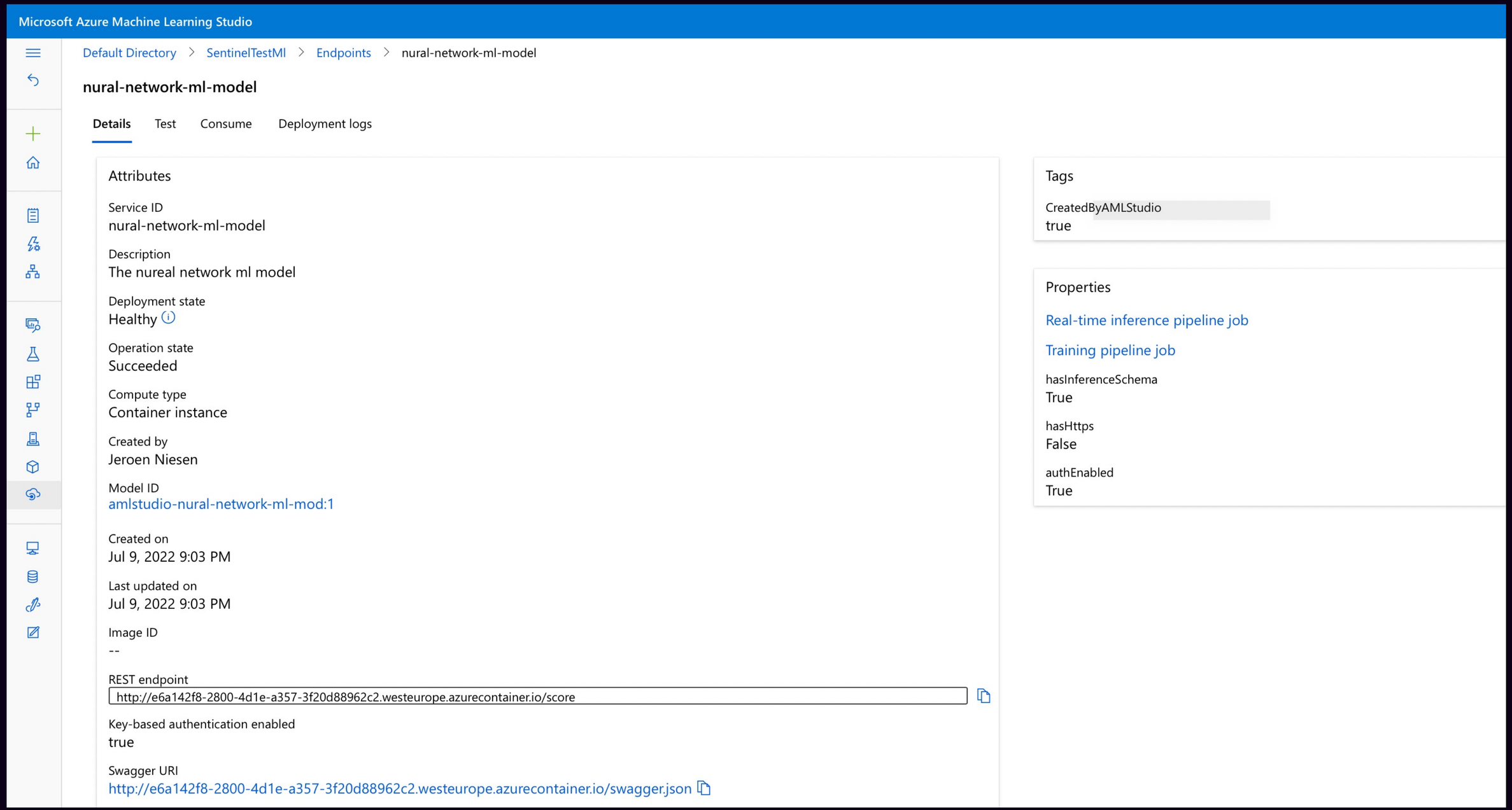
The offline approach



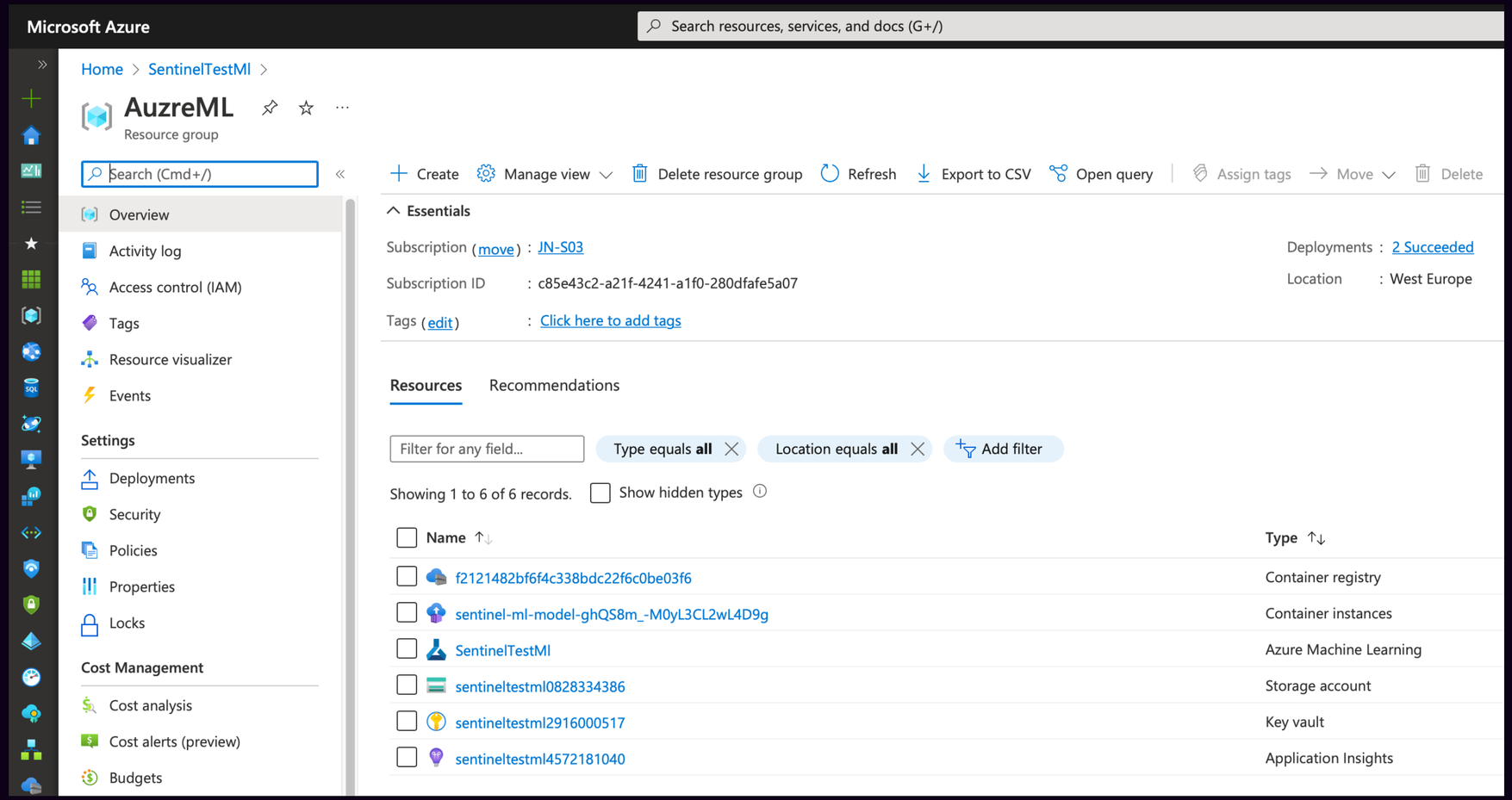
wortell

Deploying your model to Azure

- **Your model becomes a OpenAPI webservice**. Using HTTP rest calls, the webservice is capable of estimating weather a sign-in log is from a attacker or a regular user
- **Various ways of deploying**. The webservice runs in a container that can be deployed as a container instance or AKS container.


Microsoft Azure Machine Learning Studio
Default Directory > SentinelTestMl > Endpoints > nural-network-ml-model
nural-network-ml-model
Details
Test
Consume
Deployment logs
Attributes
Service ID
nural-network-ml-model
Description
The nureal network ml model
Deployment state
Healthy
Operation state
Succeeded
Compute type
Container instance
Created by
Jeroen Niesen
Model ID
amlstudio-nural-network-ml-mod:1
Created on
Jul 9, 2022 9:03 PM
Last updated on
Jul 9, 2022 9:03 PM
Image ID
--
REST endpoint
http://e6a142f8-2800-4d1e-a357-3f20d88962c2.westeurope.azurecontainer.io/score
Key-based authentication enabled
true
Swagger URI
http://e6a142f8-2800-4d1e-a357-3f20d88962c2.westeurope.azurecontainer.io/swagger.json
Tags
CreatedByAMLStudio
true
Properties
Real-time inference pipeline job
Training pipeline job
hasInferenceSchema
True
hasHttps
False
authEnabled
True

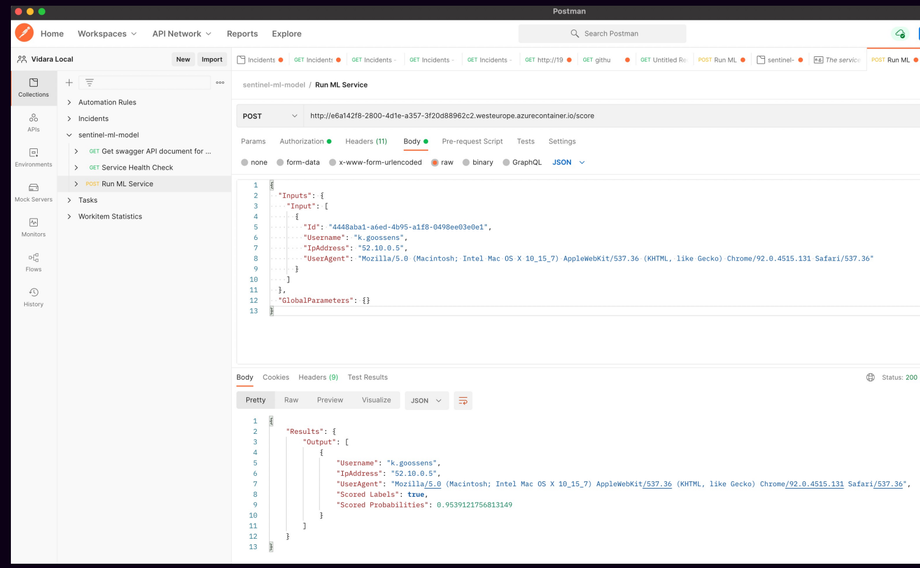
Deploying your model



wortell

Consuming your model

- **As the model is an web api, we can interact with it using postman ☺**
- **The model responds with a label and probability (number)**
- Responses close to 1 mean that the log is most likely caused by an attacker



wortell

Params | Authorization | Headers (11) | Body | Pre-request Script | Tests | Settings

none | form-data | x-www-form-urlencoded | raw | binary | GraphQL | JSON

```
{
  "Inputs": {
    "Input": [
      {
        "Id": "4448aba1-a6ed-4b95-a1f8-0498ee03e0e1",
        "Username": "k.goossens",
        "IpAddress": "52.10.0.5",
        "UserAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36"
      }
    ]
  },
  "GlobalParameters": {}
}
```

Body | Cookies | Headers (9) | Test Results

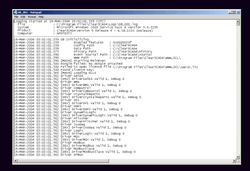
Status: 200 OK | Time: 477

Pretty | Raw | Preview | Visualize | JSON

```
{
    "Results": {
        "Output": [
            {
                "Username": "k.goossens",
                "IpAddress": "52.10.0.5",
                "UserAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/92.0.4515.131 Safari/537.36",
                "Scored Labels": true,
                "Scored Probabilities": 0.9539121756813149
            }
        ]
    }
```

The Architecture

” In-line ML classification (System tables or DCR method)”



Ingest log

Microsoft Sentinel

Export Event

Eventhub

Receives the exported log

Receive event

Logic App

Subscribed on the evenhub and used to interact with the model

Consume Model

Azure Machine Learning Workspace

Train & Deploy Model

Container Instance

Hosts the trained model

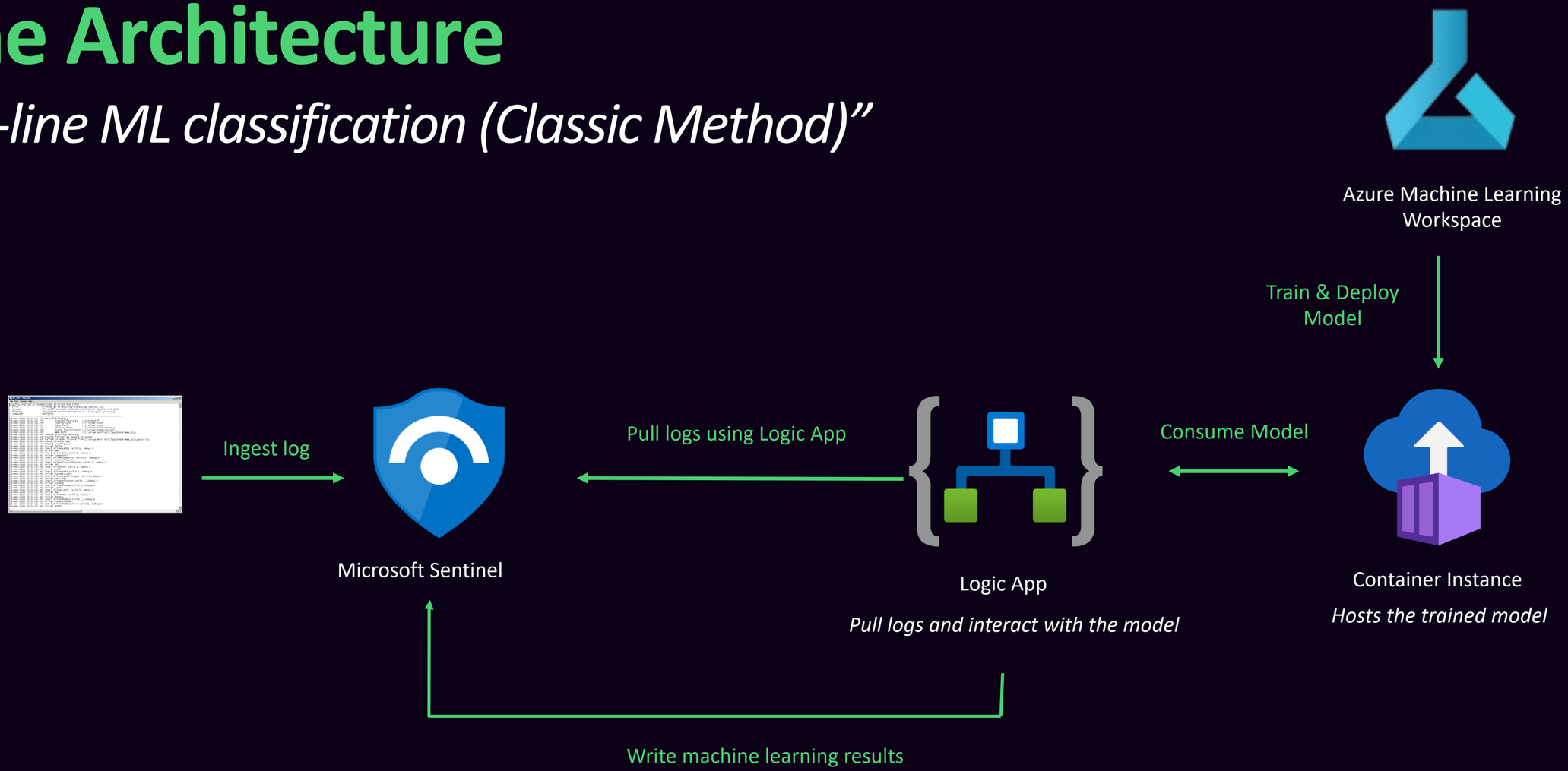
Write machine learning results

* Requires a DCR deployed table or system table, classic custom tables are not supported for exporting.

wortell

The Architecture

” In-line ML classification (Classic Method)”

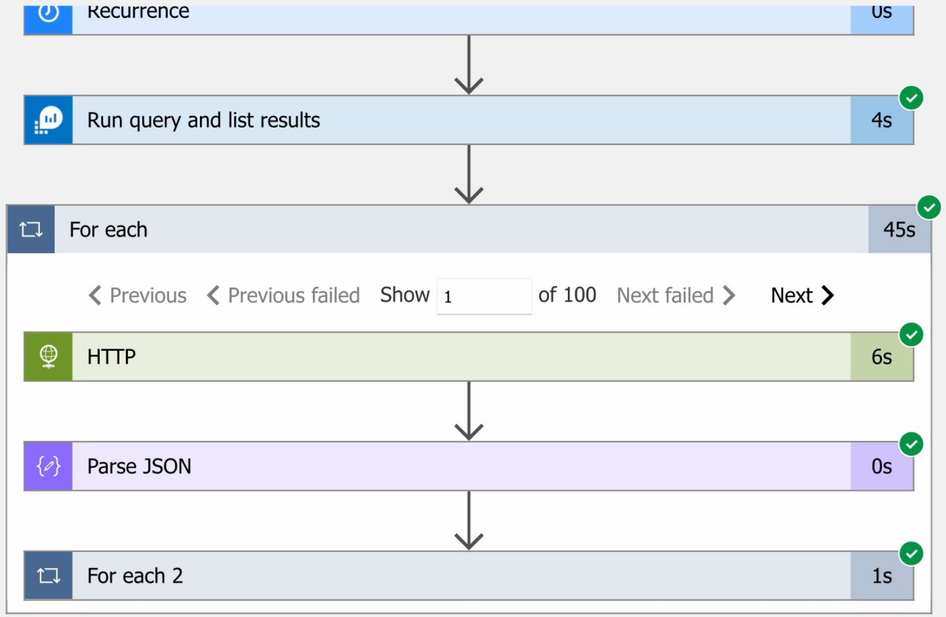


wortell

Logic Apps Designer

Save | Discard | Run Trigger | Designer | Code view | Parameters | Templates | Connectors | Help | Info

100%



New Query 1*

Feedback | Queries

JN-WEU-LA-SENTI... | Select scope | Run | Time range : **Last 24 hours** | Save | Share | New alert rule | Export | Pin to | Format query

Tables | Queries

Search

Filter | Group by: Solution

Collapse all

Favorites

You can add favorites by clicking on the ☆ icon

- **LogManagement**
- **Microsoft Sentinel**
- **Custom Logs**
  - DemoAppData_CL
  - DemoApplicationEvents_CL
  - DemoApplicationEvents_MLR...
  - DemoApplicationEvents_Nura...

```
1 DemoApplicationEvents_NuralResults_CL
2
3
```

Results | Chart

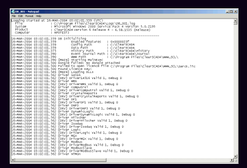
TimeGenerated [UTC]	Username_s	IPAddress	UserAgent_s	IsAnomaly_s	Probability_s	Type
7/9/2022, 8:17:32.453 PM	f.oudendorp	188.144.22.8	Mozilla/5.0 (Windows NT 10....	False	0.0151017999262965	De
7/9/2022, 8:17:33.852 PM	f.oudendorp	188.144.22.8	Mozilla/5.0 (Windows NT 10....	False	0.016199526477351	De
7/9/2022, 8:17:36.837 PM	f.oudendorp	40.80.122.5	Mozilla/5.0 (Windows NT 10....	False	0.0136263841625167	De
7/9/2022, 8:17:37.640 PM	k.goossens	188.144.22.8	Mozilla/5.0 (Windows NT 10....	False	0.00927324478185616	De
7/9/2022, 8:17:38.219 PM	f.oudendorp	188.144.22.8	Mozilla/5.0 (Windows NT 10....	False	0.0151017999262965	De
7/9/2022, 8:17:39.415 PM	f.oudendorp	40.80.122.5	Mozilla/5.0 (Windows NT 10....	False	0.0111246855552195	De
7/9/2022, 8:17:54.721 PM	m.goet	40.92.16.16	Mozilla/5.0 (Windows NT 10....	False	0.00819683592769644	De
7/9/2022, 8:20:10.784 PM	f.oudendorp	52.10.0.5	Mozilla/5.0 (Macintosh; Intel ...	True	0.95928099139574	De
7/9/2022, 8:17:39.342 PM	k.goossens	40.92.16.16	Mozilla/5.0 (Windows NT 10....	False	0.00792572906737797	De
7/9/2022, 8:20:12.170 PM	m.goet	40.92.16.16	Mozilla/5.0 (Windows NT 10....	False	0.00819683592769644	De
7/9/2022, 8:14:44.113 PM	m.goet	123.6.12.162	Mozilla/5.0 (Windows NT 10....	False	0.00926507917665554	De
7/9/2022, 8:15:11.358 PM	m.goet	40.92.16.16	Mozilla/5.0 (Windows NT 10....	False	0.00733125866285009	De
7/9/2022, 8:15:12.649 PM	j.niesen	40.92.16.16	Mozilla/5.0 (Windows NT 10....	False	0.00789125576621194	De
7/9/2022, 8:15:20.401 PM	m.goet	188.144.22.8	Mozilla/5.0 (Windows NT 10....	False	0.00843068100160418	De
7/9/2022, 8:15:34.587 PM	k.goossens	40.92.16.16	Mozilla/5.0 (Windows NT 10....	False	0.00562408336413199	De
7/9/2022, 8:15:46.504 PM	f.oudendorp	188.144.22.8	Mozilla/5.0 (Windows NT 10....	False	0.0151017999262965	De

Columns

0s 469ms | Display time (UTC+00:00) | Query details | 29 - 45 of 175

The Architecture

”Interaction using Jupyter notebooks”



Ingest log

Microsoft Sentinel

Pull logs

Jupyter notebook

Consume Model

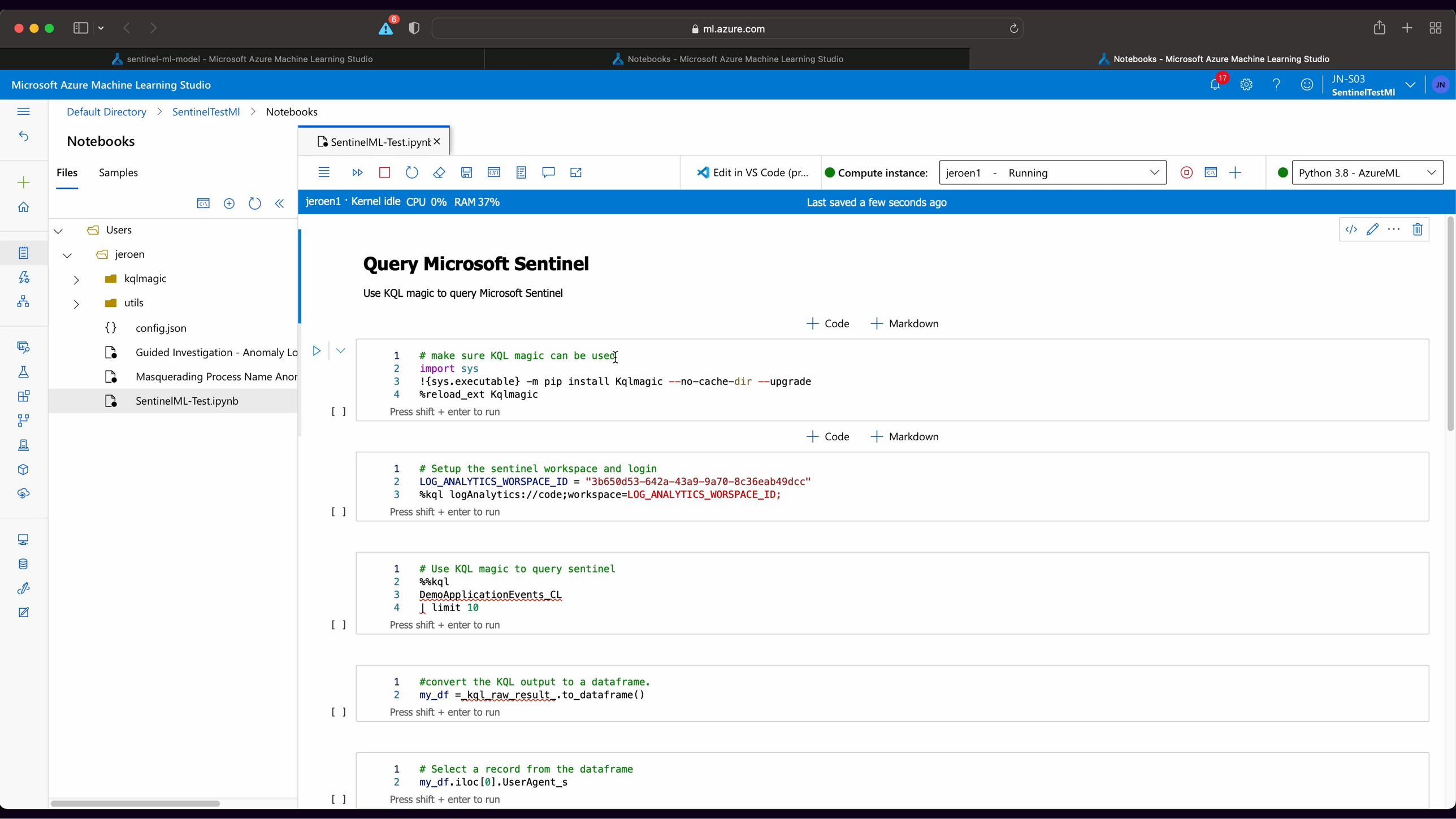
Azure Machine Learning Workspace

Train & Deploy Model

Container Instance

Hosts the trained model

wortell


ml.azure.com
sentinel-ml-model - Microsoft Azure Machine Learning Studio
Notebooks - Microsoft Azure Machine Learning Studio
Notebooks - Microsoft Azure Machine Learning Studio
Microsoft Azure Machine Learning Studio
JN-S03
SentinelTestMl
Default Directory > SentinelTestMl > Notebooks
Notebooks
Files
Samples
Users
jeroen
kqlmagic
utils
config.json
Guided Investigation - Anomaly Lo
Masquerading Process Name Anor
SentinelML-Test.ipynb
SentinelML-Test.ipynb
Edit in VS Code (pr...
Compute instance:
jeroen1 - Running
Python 3.8 - AzureML
jeroen1 · Kernel idle CPU 0% RAM 37%
Last saved a few seconds ago
Query Microsoft Sentinel
Use KQL magic to query Microsoft Sentinel
+ Code
+ Markdown
# make sure KQL magic can be used
import sys
!{sys.executable} -m pip install Kqlmagic --no-cache-dir --upgrade
%reload_ext Kqlmagic
Press shift + enter to run
+ Code
+ Markdown
# Setup the sentinel workspace and login
LOG_ANALYTICS_WORSPACE_ID = "3b650d53-642a-43a9-9a70-8c36eab49dcc"
%kql logAnalytics://code;workspace=LOG_ANALYTICS_WORSPACE_ID;
Press shift + enter to run
Use KQL magic to query sentinel
%%kql
DemoApplicationEvents_CL
| limit 10
Press shift + enter to run
#convert the KQL output to a dataframe.
my_df = _kql_raw_result_.to_dataframe()
Press shift + enter to run
Select a record from the dataframe
my_df.iloc[0].UserAgent_s
Press shift + enter to run

Data lakehouse platforms

Azure Machine Learning Workspace

Azure Synapse Analytics

wortell

Steps to create a model

1. Understand data
2. Have access to the data
3. Transform data so it can be used in machine learning
4. Split the dataset into a training dataset and validation dataset
5. Select your machine learning model
6. Train your ML model
7. Validate your ML model
8. Use your model

wortell

Mario's adventure is over. Mario, the Princess, Yoshi, and his friends are going to take a vacation.

Conclusion

- Knowing your data is key
- Having access to the values that you try to predict is essential
- Experimenting with different methods can pay off

wortell


THANK YOU
wortell