

Machine learning project

Network Intrusion Detection system

Problem statement No.40 – Network Intrusion Detection

The Challenge:

Create a robust network intrusion detection system (NIDS) using machine learning. The system should be capable of analysing network traffic data to identify and classify various types of cyber-attacks (e.g., DoS, Probe, R2L, U2R) and distinguish them from normal network activity. The goal is to build a model that can effectively secure communication networks by providing an early warning of malicious activities.

About Dataset:

Background

The dataset to be audited was provided which consists of a wide variety of intrusions simulated in a military network environment. It created an environment to acquire raw TCP/IP dump data for a network by simulating a typical US Air Force LAN. The LAN was focused like a real environment and blasted with multiple attacks. A connection is a sequence of TCP packets starting and ending at some time duration between which data flows to and from a source IP address to a target IP address under some well-defined protocol. Also, each connection is labelled as either normal or as an attack with exactly one specific attack type. Each connection record consists of about 100 bytes.

For each TCP/IP connection, 41 quantitative and qualitative features are obtained from normal and attack data (3 qualitative and 38 quantitative features). The class variable has two categories:

- Normal
- Anomalous

[Kaggle dataset link](#) - Used in this representation – this will generate a model that only give prediction of class(normal or anomaly)

Technology Used - IBM cloud lite services.

[GitHub repo link](#)

For upgraded model:

This model will classify network attacks like **DoS, Probe, R2L, U2R**

First, quick attack type breakdown:

Attack Type	Description
DoS (Denial of Service)	Overwhelm system/network to make services unavailable
Probe	Scanning/snooping to gather network info
R2L (Remote to Local)	Remote attacker tries to gain local access
U2R (User to Root)	Attacker with user access tries to escalate to root

1. DoS Attacks

These often involve **large volume** or **abnormal traffic**.

- `src_bytes`, `dst_bytes`: Sudden large/small byte transfer.
- `count`, `srv_count`: High numbers indicate a burst of connections.
- `error_rate`, `srv_error_rate`: High error rates (typical in SYN flood).
- `dst_host_srv_count`, `dst_host_same_srv_rate`: Repeated connection to same service on host.

Red flags for DoS:

High `count`, `error_rate`, `dst_host_srv_count` with low `diff_srv_rate`.

2. Probe Attacks

These are scans/reconnaissance, often *low payload, high variety*.

- `count`, `srv_count`: High, but over varied services.
- `same_srv_rate`, `diff_srv_rate`, `srv_diff_host_rate`: Show diversity of target services/hosts.
- `dst_host_diff_srv_rate`: High for scan-like behavior.

Red flags for Probe:

High `diff_srv_rate`, high `dst_host_diff_srv_rate`, moderate `count`.

3. R2L Attacks

Low-volume but abnormal login patterns.

- `logged_in`: Often 0 (unsuccessful attempts).
- `num_failed_logins`: High.
- `is_guest_login`: 1 (suspicious).
- `hot`, `num_access_files`, `num_compromised`: Suspicious if increased.
- `service`: Often mail, ftp, telnet, or HTTP.

Red flags for R2L:

`num_failed_logins` > 0, `logged_in` = 0, maybe `is_guest_login` = 1.

4. U2R Attacks

Attempting to escalate privileges. Very subtle.

- `num_compromised`, `root_shell`, `su_attempted`, `num_root`: Big signs.
- `hot`, `num_file_creations`, `num_shells`: High if privilege escalation occurs.

- Very **low** src_bytes/dst_bytes, often **single connection**.

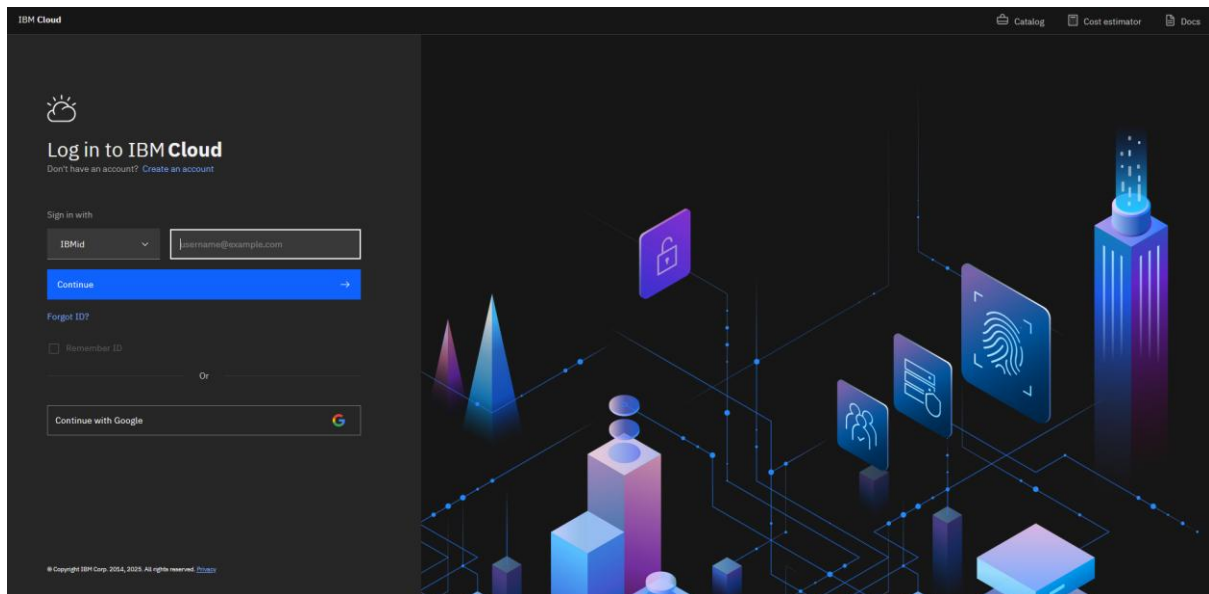
Red flags for U2R:

root_shell = 1, num_file_creations > 0, su_attempted = 1, low traffic volume.

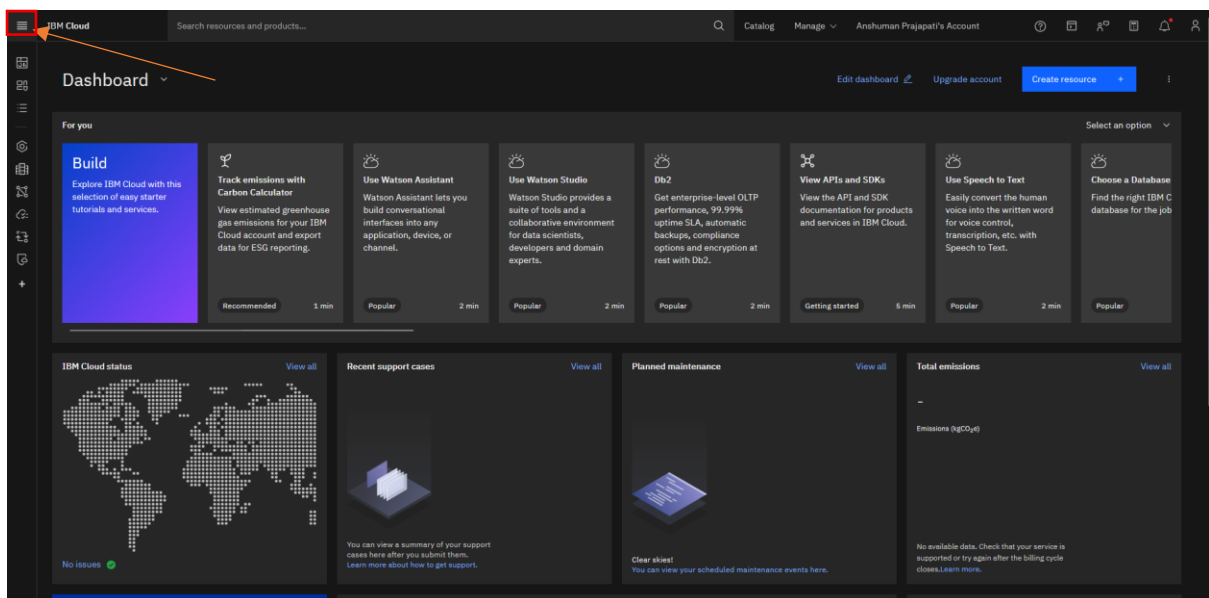
DATA sets used for this upgraded model:

[-New data set for training of model](#)

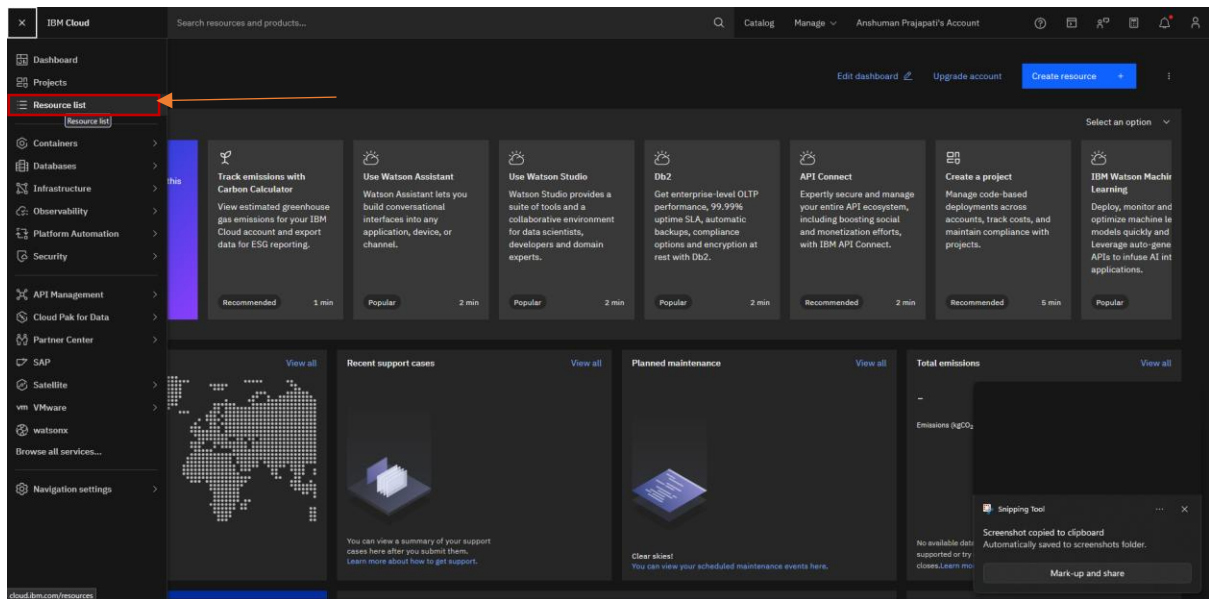
[-New data set for testing of model](#)



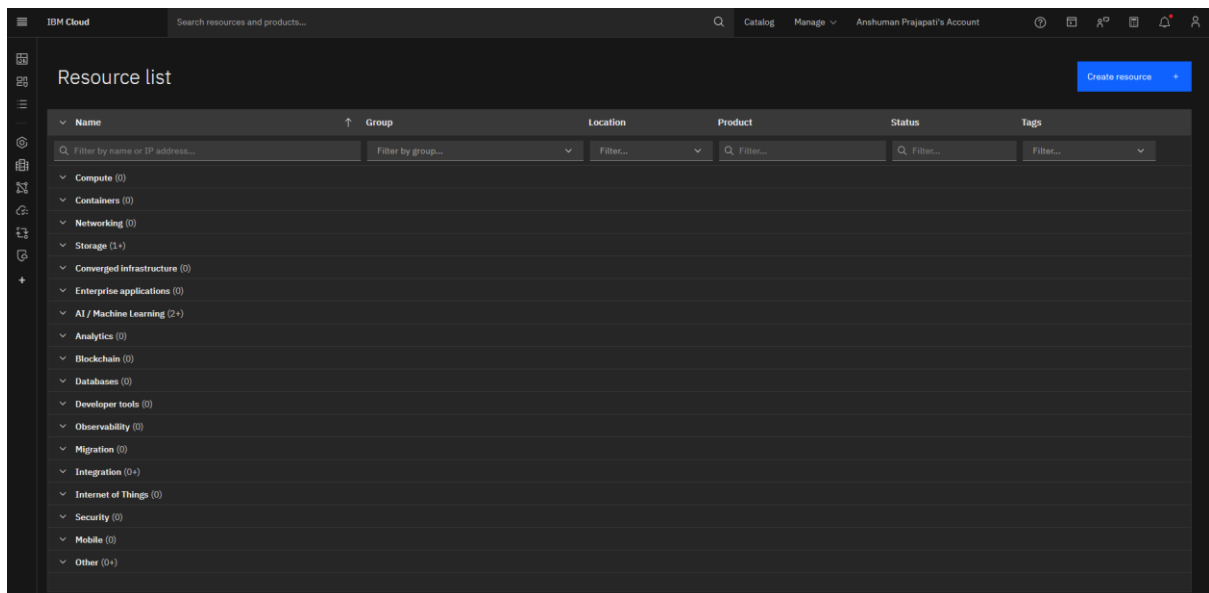
1. Login to [IBM cloud](#) with your credentials.



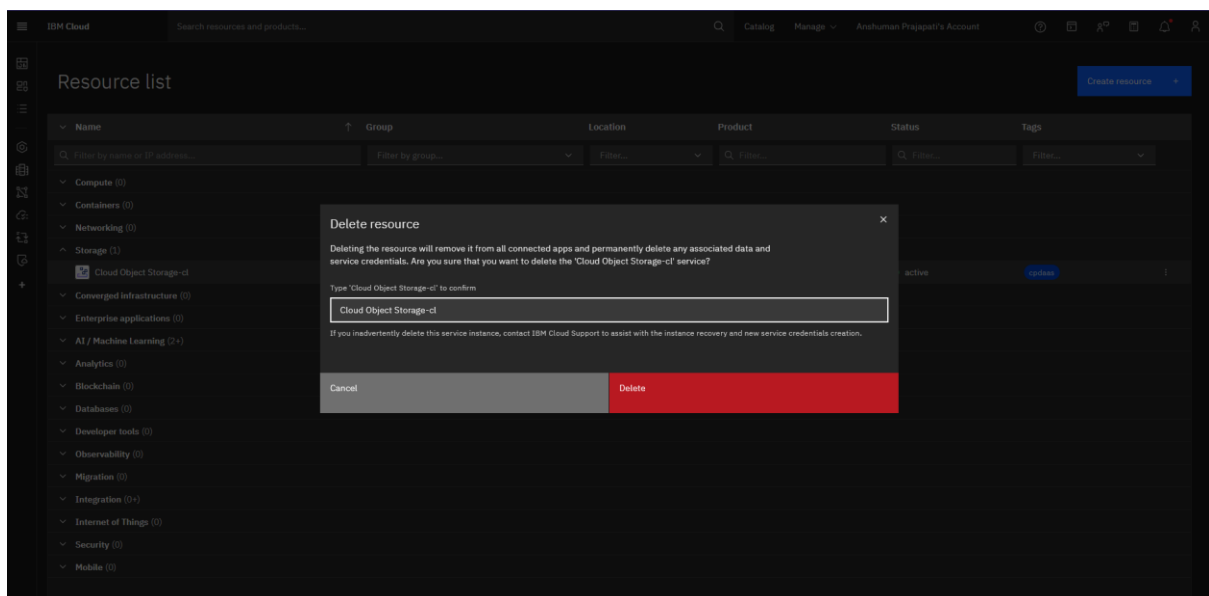
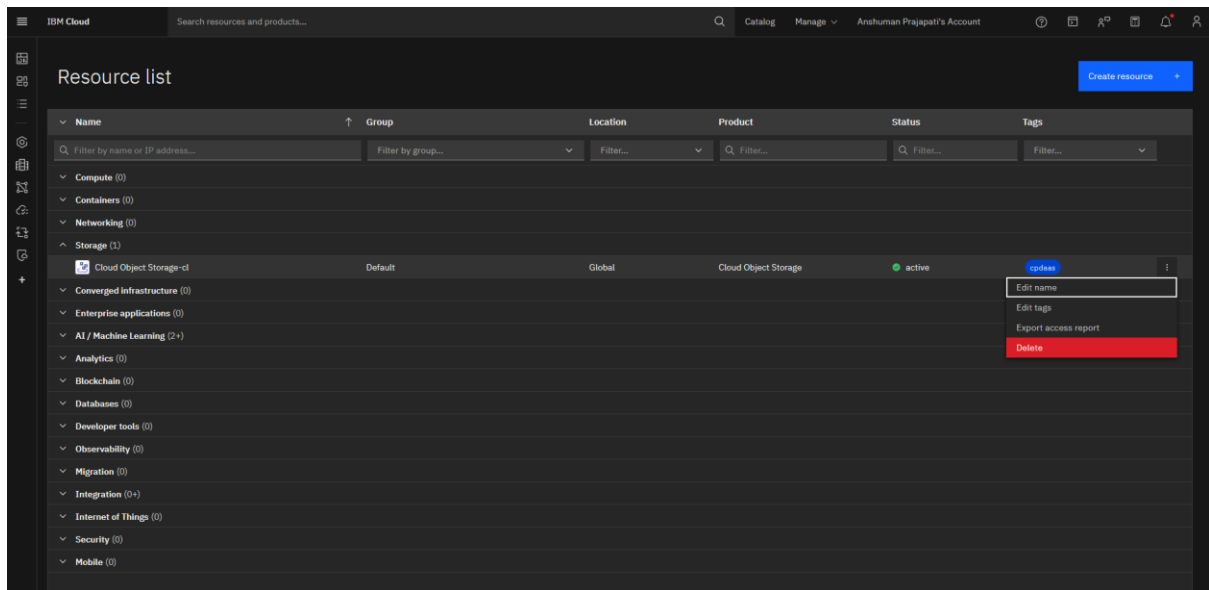
2. This is homepage of IBM cloud now press on top left corner of this page.



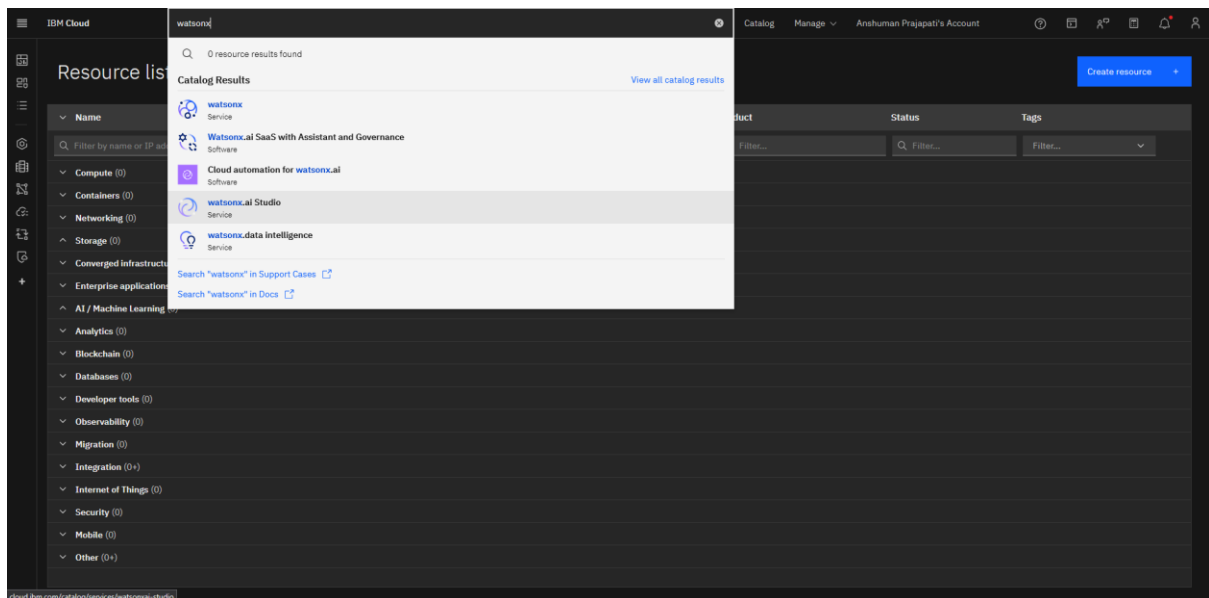
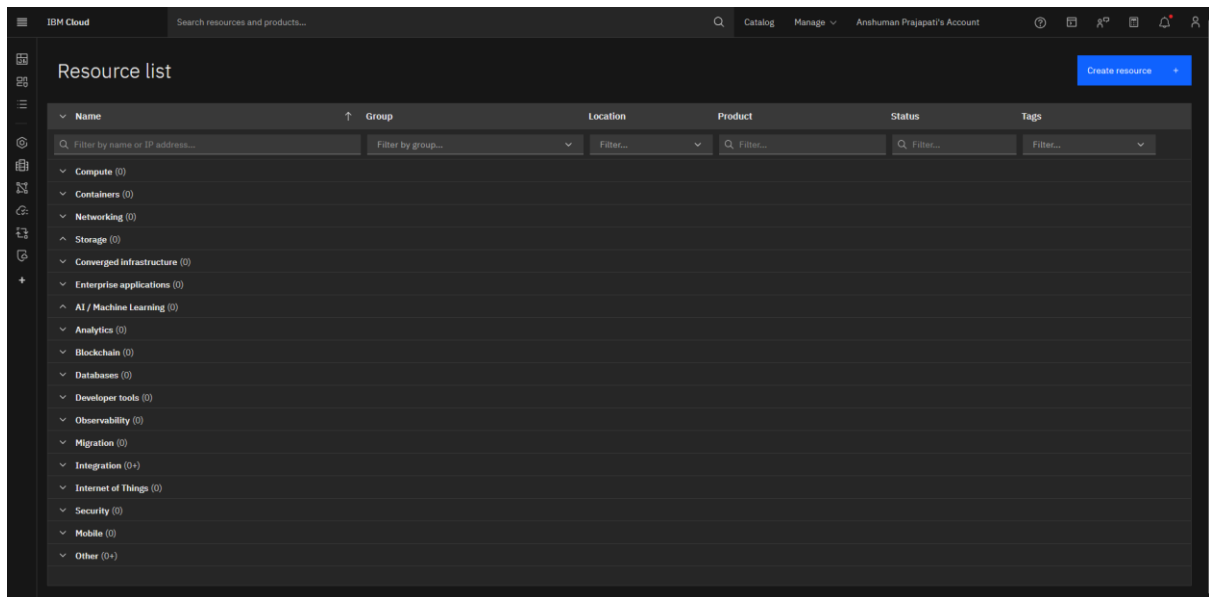
3. Press on Resource list



4. In this page if any resource is in use delete it and make it free.



- Repeat this for all the occupied resources and make sure no resources are in use.



6. Now click search for watsonx.ai Studio in search bar and press enter.

watsonx.ai Studio
(Formerly known as Watson Studio) Develop powerful AI solutions with an integrated collaborative studio and industry-standard APIs and SDKs.

Create | About

Type: Service
Provider: IBM
Last updated: 05/06/2025
Category: AI / Machine Learning
Compliance: HIPAA Enabled, IAM-enabled
Location: Sydney (au-syd), Frankfurt (eu-de), London (eu-gb), Tokyo (jp-tok), Dallas (us-south), Toronto (ca-tor)
Related links: Docs, Terms

Select a location: **Sydney (au-syd)**

Select a pricing plan
Prices shown are for country or location: [United States](#)

Plan	Features and capabilities	Pricing
Lite	<ul style="list-style-type: none"> 1 authorized user 10 capacity units monthly limit Environment = # of capacity units required per hour • 1 vCPU + 4 GB RAM = 0.5 • 2 vCPU + 8 GB RAM = 1 • 4 vCPU + 16 GB RAM = 2 • Decision Optimization + Watson NLP = Environment + 5 • Synthetic Data Generator, 2 vCPU + 8 GB RAM = 7 (requires watsonx.ai Runtime) 	Free
Professional	<ul style="list-style-type: none"> Unlimited collaborators Unlimited elastic compute environments Environment = # of capacity units required per hour • 1 vCPU + 4 GB RAM = 0.5 • 2 vCPU + 8 GB RAM = 1 • 4 vCPU + 16 GB RAM = 2 	\$1.02 USD/Capacity Unit-Hour

The Lite plan offers most watsonx.ai Studio data science and AI features with usage restrictions.
Lite plan services are deleted after 30 days of inactivity.

Summary
watsonx.ai Studio
Location: Sydney (au-syd)
Plan: Lite
Service name: watsonx.ai Studio-07
Resource group: Default

☒ I have read and agree to the following license agreements:
[Terms](#)

Create
Add to estimate

7. Now create with lite plan and and make sure to check on license agreements.

watsonx.ai Studio-07 Add tags

Manage
Plan

watsonx.ai Studio in Cloud Pak for Data and watsonx

Build and deploy machine learning models on either platform.
Work with foundation models on watsonx as a Service.

Launch in **Launch in**

IBM watsonx.ai Studio in Cloud Pak for Data and watsonx

IBM Cloud Pak for Data, watsonx
Unifying platforms

IBM Cloud
Base cloud infrastructure

IBM watsonx.ai Studio is part of IBM Cloud Pak for Data and watsonx, and serves as the AI capability of the data fabric architecture.

Helpful links

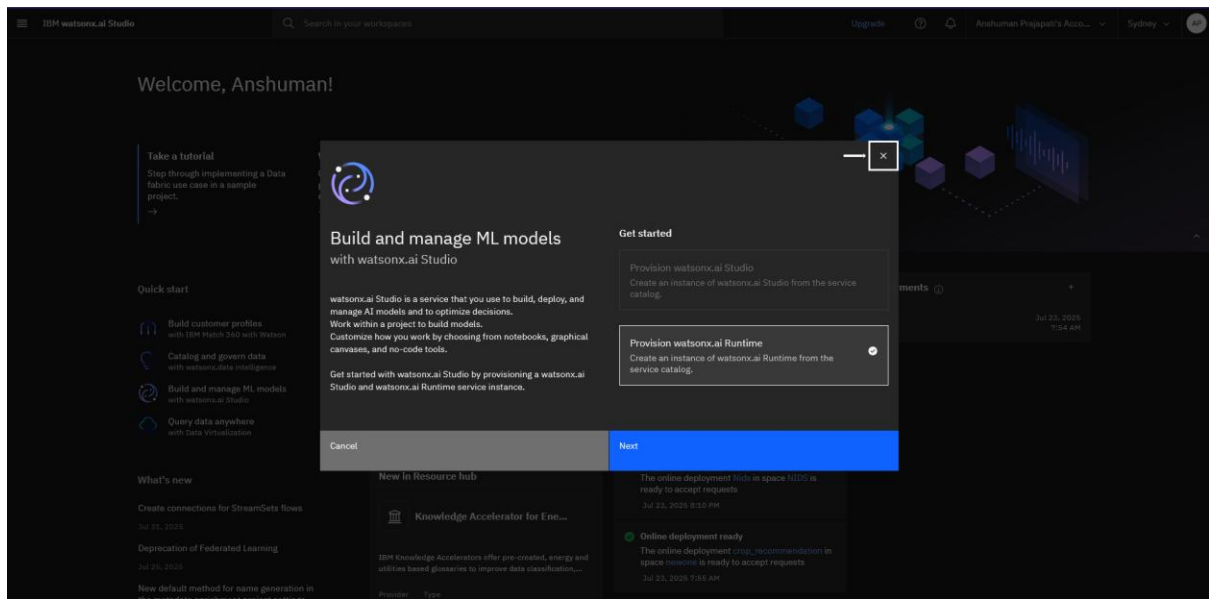
Documentation
Learn about tools, features, and how to perform a wide variety of Data and AI tasks.
[Cloud Pak for Data](#) → [watsonx](#) →

Learning path
Start a step-by-step tutorial to get up and running quickly.
[Cloud Pak for Data](#) → [watsonx](#) →

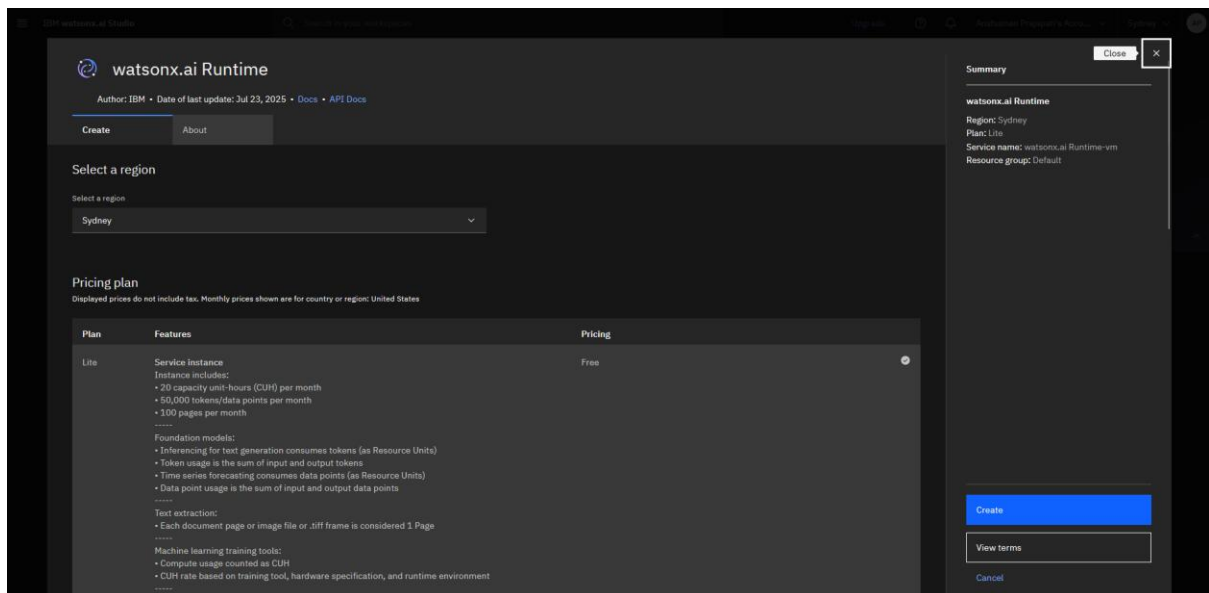
Videos
Watch videos to learn about watsonx.ai Studio.
[Cloud Pak for Data](#) → [watsonx](#) →

How to use watsonx.ai Studio

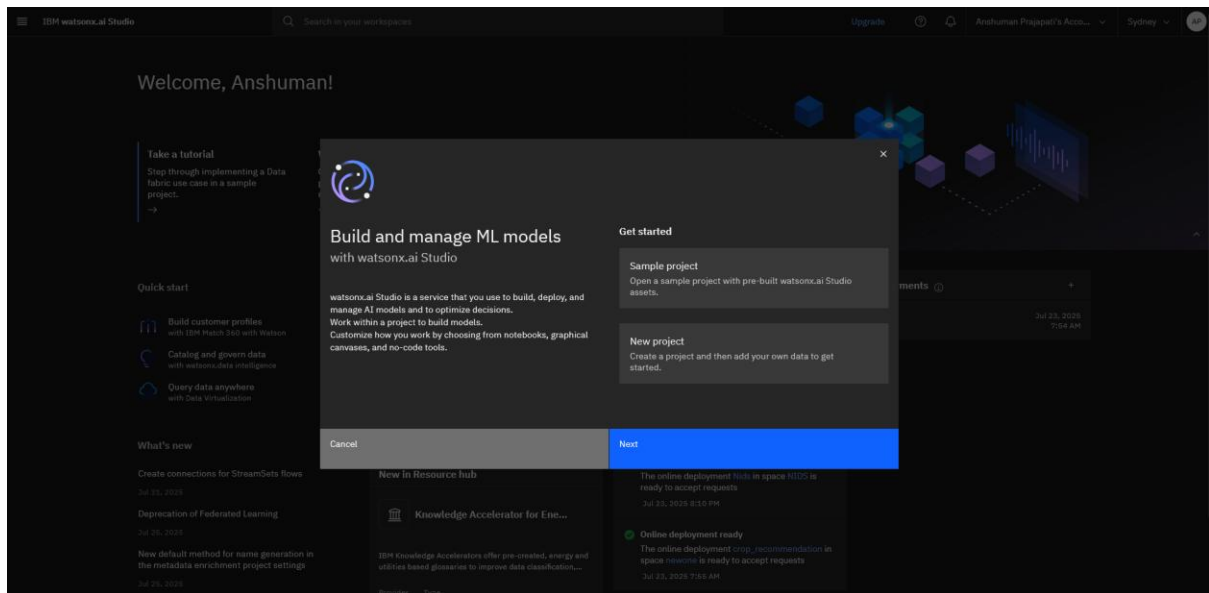
8. Now click on launch in.



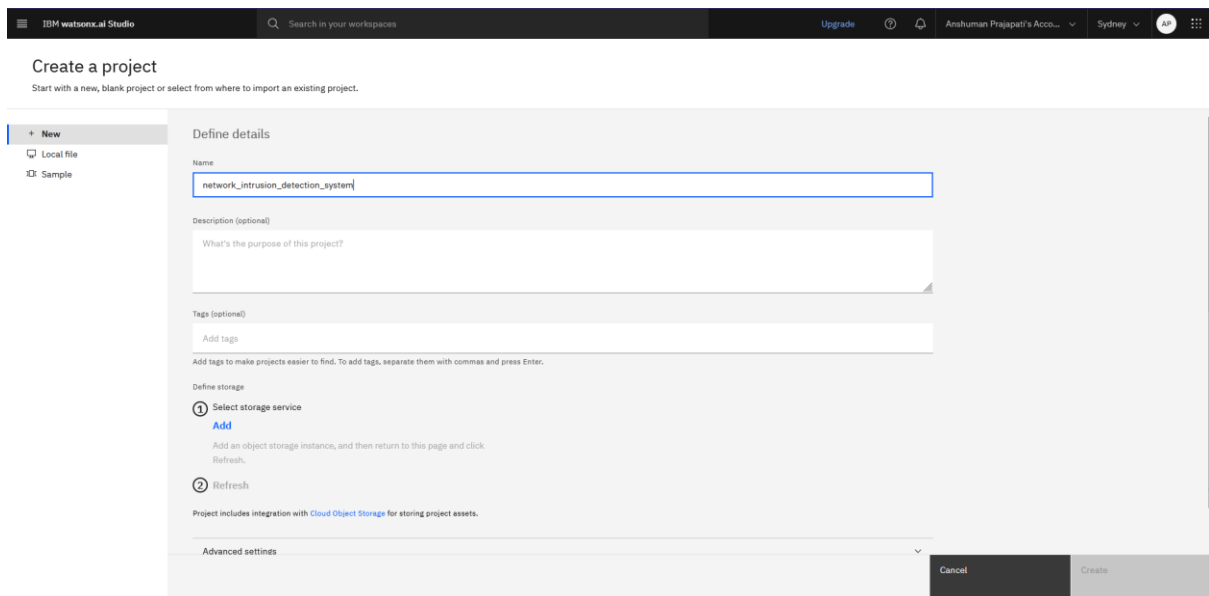
9. Now select provision watsonx.ai Runtime and press next.



10. Press create.



11. Select new project and press next.



12. Give the project a name.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade ⓘ 🔔

Anshuman Prajapati's Acco... Sydney

Create a project

Start with a new, blank project or select from where to import an existing project.

+ New

Local file

Sample

Define details

Name

network_intrusion_detection_system

Description (optional)

What's the purpose of this project?

Tags (optional)

Add tags

Add tags to make projects easier to find. To add tags, separate them with commas and press Enter.

Define storage

1 Select storage service

Add

Add an object storage instance, and then return to this page and click Refresh.

2 Refresh

Project includes integration with [Cloud Object Storage](#) for storing project assets.

Advanced settings

Cancel

Create

13. Now add storage to it.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade ⓘ 🔔

Anshuman Prajapati's Acco... Sydney

Services catalog /

Cloud Object Storage

Author: IBM • Date of last update: Apr 15, 2025 • Docs • API Docs

Create

About

Pricing plan

Displayed prices do not include tax. Monthly prices shown are for country or region: United States

Plan	Features	Pricing
One-Rate	One-Rate Plan is a Pay-as-You-Go option with a single, flat monthly rate (\$/GB) that includes storage, API operations, retrieval, and outbound bandwidth—making it ideal for high-activity workloads with frequent access and data transfer, such as analytics, media, and web apps. The plan includes built-in allowances that scale with stored capacity and offers automatic volume discounts as usage grows	
Lite(deprecated)	<div> <div>Lite plan instance is free to use for Storage capacity up to 25 GB per month. Lite plan instance is used for trial, and can be easily upgraded to Standard plan for unlimited scalability and full functionality.</div> <div>None</div> <div>Lite plan services are deleted after 30 days of inactivity.</div> </div>	Free
Standard	<div>Standard Plan is a flexible Pay-as-You-Go option with no minimum fee—ideal for workloads with large storage needs but low or infrequent access and outbound traffic. It includes a Free Tier with 5GB of Smart Tier storage for 12 months. Charges are based on actual usage, with separate billing for storage, outbound bandwidth, API operations, and data retrieval. Multiple storage classes help you optimize costs based on how often data is accessed.</div> <div>Free Tier allowance:</div> <div>Storage up to 5GB/month</div> <div>Up to 2000 Class A requests/month</div> <div>Up to 20,000 Class B requests/month</div> <div>Up to 10GB/month of data retrieval</div> <div>Up to 5GB/month of egress</div> <div>Applies to aggregate total across all smart tier buckets in your account</div>	

Summary

Cloud Object Storage

Region: Global

Plan: Lite(deprecated)

Service name: Cloud Object Storage-zg

Resource group: Default

Create

View terms

Cancel

14. Create a object storge in lite plan.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Acco...

Sydney

AP

Create a project

Start with a new, blank project or select from where to import an existing project.

+ New

Local file

Sample

Define details

Name

network_intrusion_detection_system

Description (optional)

What's the purpose of this project?

Tags (optional)

Add tags

Add tags to make projects easier to find. To add tags, separate them with commas and press Enter.

Define storage

1 Select storage service

Add

Add an object storage instance, and then return to this page and click Refresh.

2 Refresh

Project includes integration with [Cloud Object Storage](#) for storing project assets.

Advanced settings

Cancel

Create

15. Press refresh to show up the storage.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Acco...

Sydney

AP

Create a project

Start with a new, blank project or select from where to import an existing project.

+ New

Local file

Sample

Define details

Name

network_intrusion_detection_system

Description (optional)

What's the purpose of this project?

Tags (optional)

Add tags

Add tags to make projects easier to find. To add tags, separate them with commas and press Enter.

Storage

Cloud Object Storage-ng

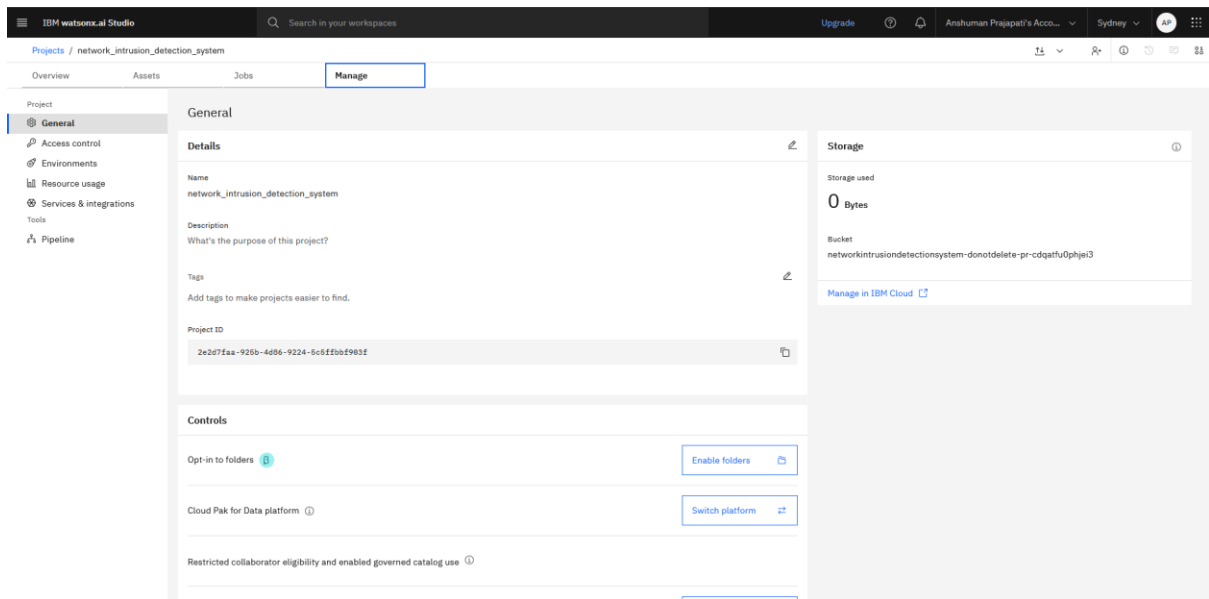
Project includes integration with [Cloud Object Storage](#) for storing project assets.

Advanced settings

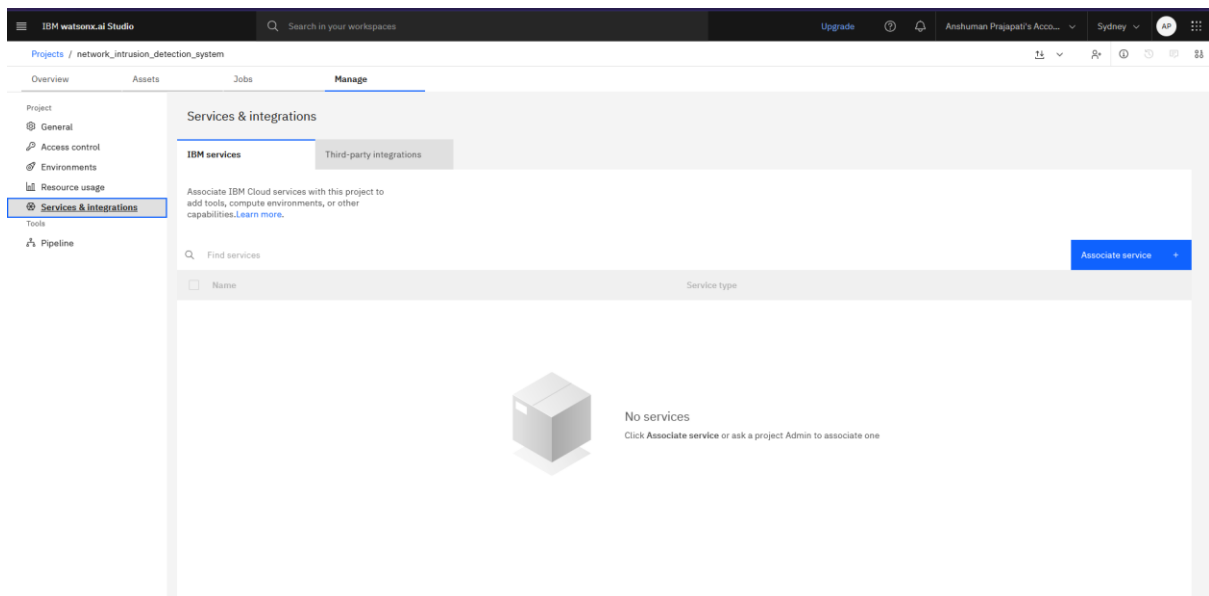
Cancel

Create

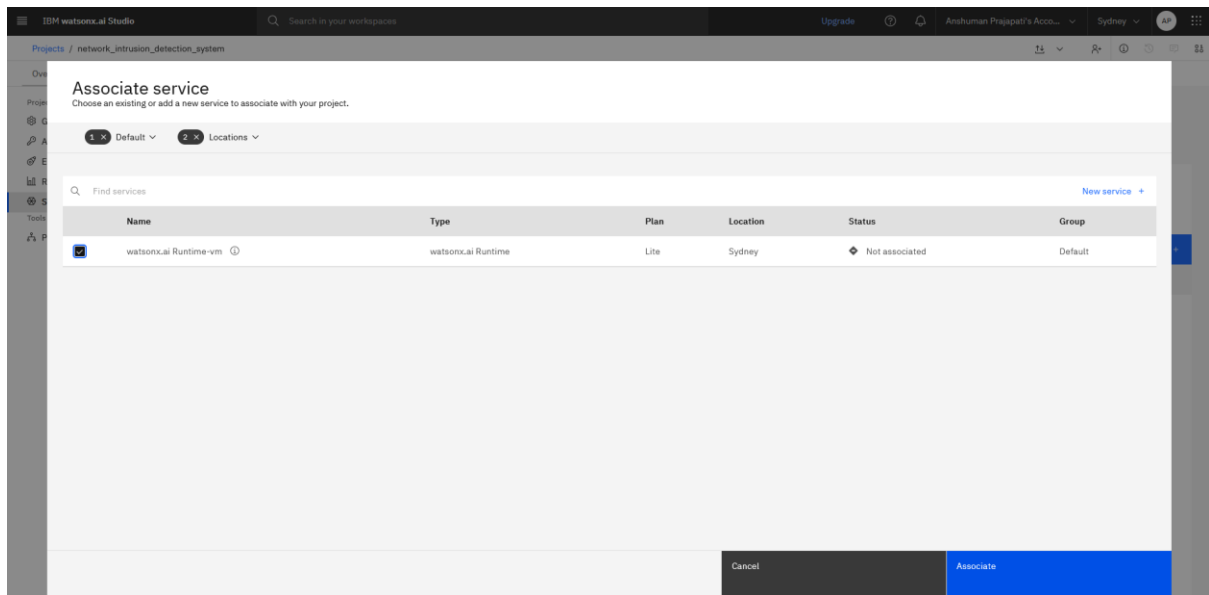
16. Press create.



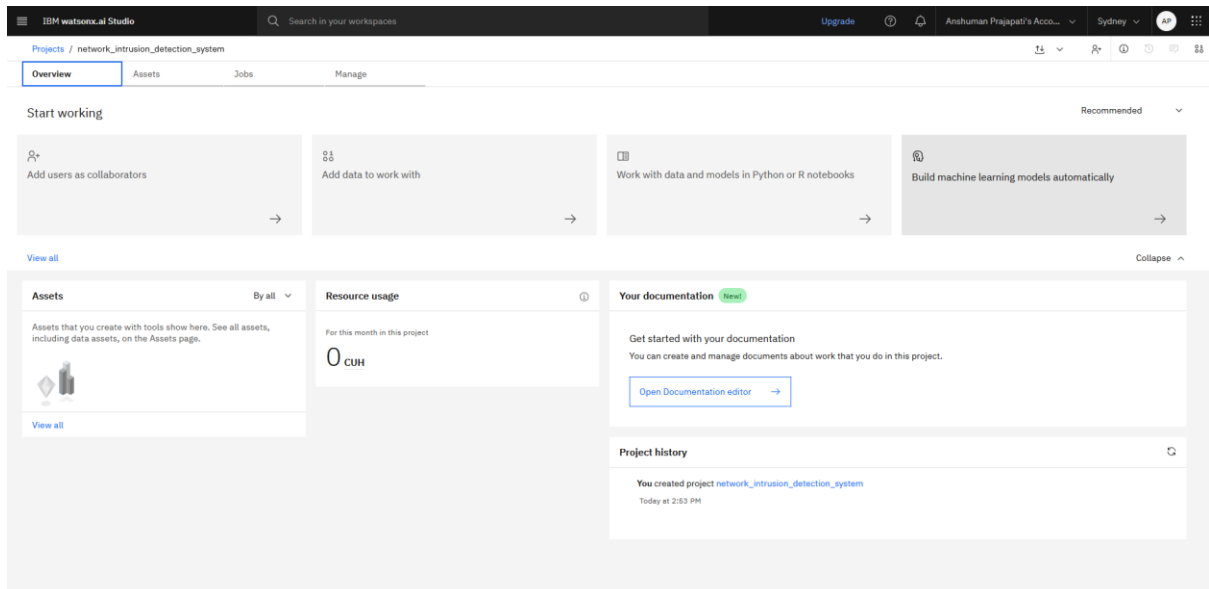
17. Now on manage tab go to services and integration.



18. Now click on associate service.



19. Check on watsonx.ai Runtime service and associate it.



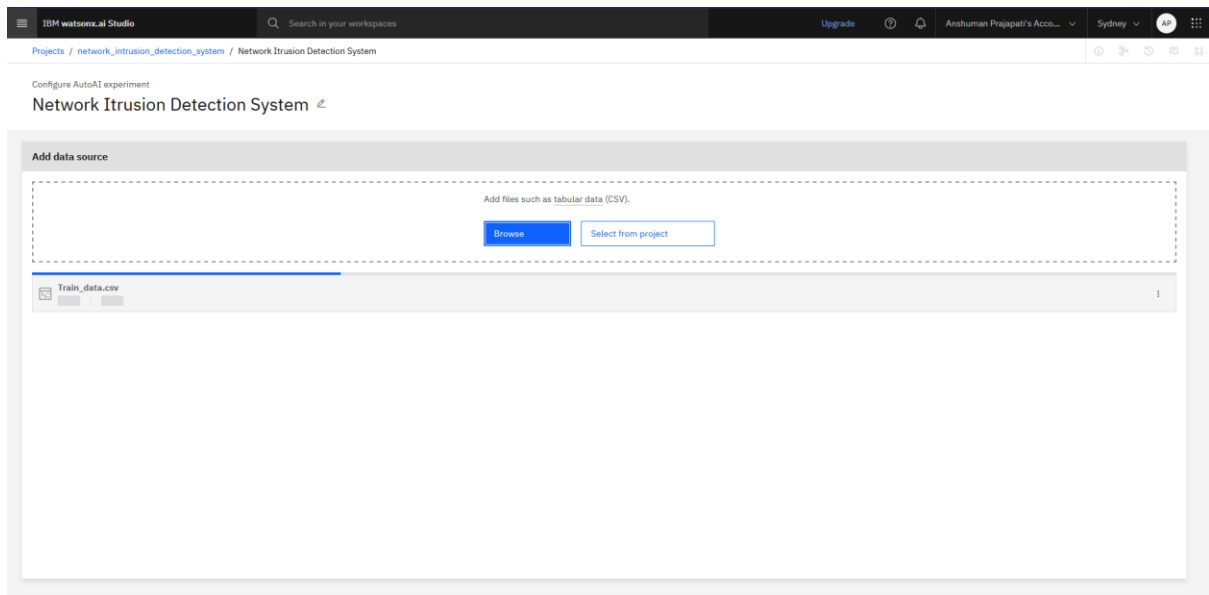
20. Now comeback to overview tab and select **Build machine learning models automatically**

The screenshot shows the 'Build machine learning models automatically' form in IBM watsonx.ai Studio. The form is divided into two main sections: 'Define details' and 'Define configuration'. In the 'Define details' section, the 'Name' field is filled with 'Network Intrusion Detection System'. The 'Description (optional)' field contains the text 'What's the purpose of this AutoAI experiment?'. The 'Tags (optional)' field is empty. In the 'Define configuration' section, the 'Name' field is filled with 'watsonx.ai Runtime-vm'. The 'Environment definition' field is filled with 'Large: 8 CPU and 32 GB RAM'. A note below states: 'This environment definition consumes 20 capacity units per hour for training. For details, see [watsonx.ai Runtime plans](#).' At the bottom of the form, there are three buttons: 'Cancel', 'Back', and 'Create'.

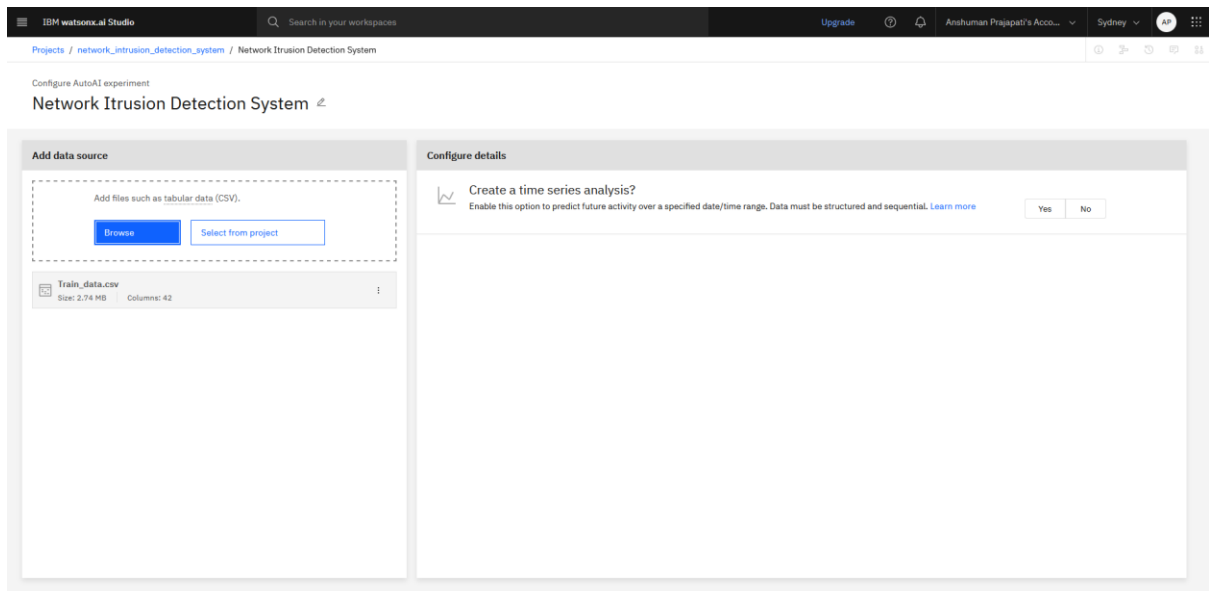
21. Give the experiment a name and click on create.

The screenshot shows the 'Add data source' form in IBM watsonx.ai Studio. The form is titled 'Configure AutoAI experiment' and 'Network Intrusion Detection System'. The main area is a large dashed box containing an illustration of a person standing next to a computer monitor. To the right of the illustration, the text reads: 'Drop data files here or browse for files to upload'. Below this, it says: 'Add files such as tabular data (CSV)'. There are two buttons: 'Browse' and 'Select from project'.

22. Now browse the file to upload (i.e. the .csv file to train the model).



23. Wait for the file to upload.



24. Now after the file gets completely uploaded it will ask for time series analysis select **no**.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Account

Sydney

Projects / network_intrusion_detection_system / Network Intrusion Detection System

Configure AutoAI experiment

Network Intrusion Detection System

Autosaved: 3:15:35 pm

Add data source

Add files such as tabular data (CSV).

[Browse](#) [Select from project](#)

Train_data.csv
Size: 2.74 MB Columns: 42

Configure details

Create a time series analysis?
Enable this option to predict future activity over a specified data/time range. Data must be structured and sequential. [Learn more](#)

What do you want to predict?
Prediction column:

Prediction column: class CUH remaining: 12:14 CUH

PREDICTION TYPE
Multiclass Classification

OPTIMIZED FOR
Accuracy & run time

Experiment settings [Run experiment](#)

25. Select the prediction type (in this case its class) and make sure the prediction type is multiclass classification(as prediction column contains multiple distinct categories) and run experiment.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Account

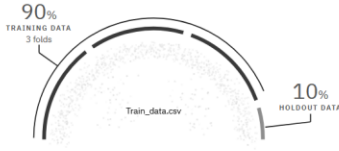
Sydney

Projects / network_intrusion_detection_system / Network Intrusion Detection System

Experiment summary Pipeline comparison

Rank by: Accuracy (Optimized) Cross validation score

Relationship map [ⓘ]
Prediction column: class



90% TRAINING DATA 3 folds

10% HOLDOUT DATA

Train_data.csv

Progress map

Swap view

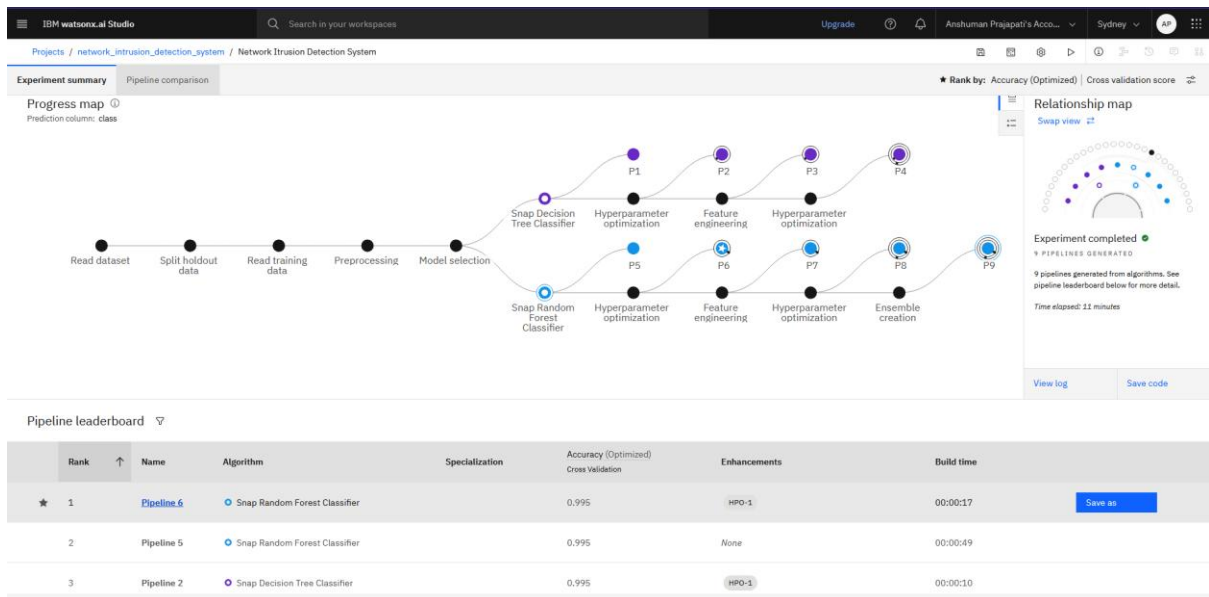
Preprocessing
TRAIN_DATA.CSV
Setting default preprocessor parameters
Time elapsed: 2 minutes

[View log](#) [Save code](#)

Pipeline leaderboard

Rank	↑	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
...						

26. It will take some time to perform the experiment.



27. Now after the experiment is completed we can see the best performing algorithm is Pipeline 6 with enhancement of hyperparameter optimization now click on save as.

The 'Save as' dialog is open, showing options to save the pipeline as a 'Model' or a 'Notebook'. The 'Model' option is selected. The 'Define details' section on the right contains the following information:

- Name:** P6 - Snap Random Forest Classifier: Network Intrusion Detection System
- Description (optional):** Model description
- Tags:** Add tags to make assets easier to find. (Add a tag button)

At the bottom right, there are 'Cancel' and 'Create' buttons.

28. Press on create.

Experiment summary Pipeline comparison

Progress map Prediction column: class

★ Rank by:

Experiment completed 9 PIPELINES GENERATED
9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 11 minutes

View log Save code

Pipeline leaderboard

Rank	↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★ 1		Pipeline 6	Snap Random Forest Classifier		0.995	HPO-1	00:00:17
2		Pipeline 5	Snap Random Forest Classifier		0.995	None	00:00:49
3		Pipeline 2	Snap Decision Tree Classifier		0.995	HPO-1	00:00:10

29. After the model is saved click on view in project.

Input (1)

Column	↑	Type
count		double
diff_srv_rate		double
dst_bytes		double
dst_host_count		double
dst_host_diff_srv_rate		double
dst_host_error_rate		double
dst_host_same_src_port_rate		double
dst_host_same_srv_rate		double

Promote to space

About this asset

Name
P6 - Snap Random Forest Classifier: Network Intrusion Detection System

Description
No description provided.

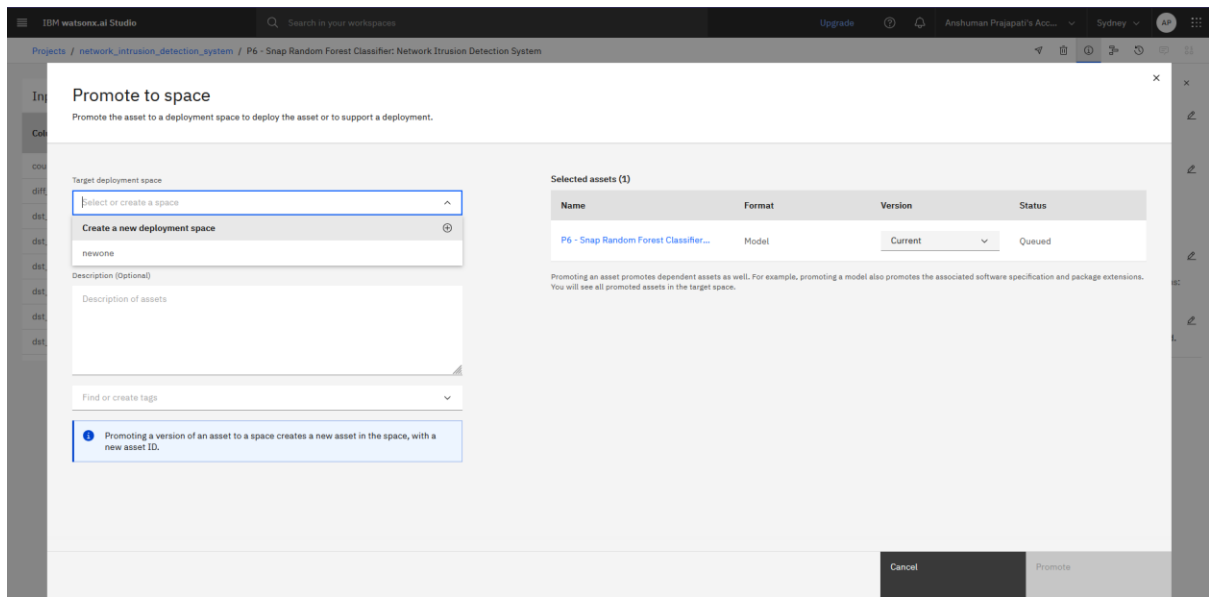
Asset Details
Type: wml-hybrid_0.1
Model ID: c7644f35-5826-4c...
Software specification: hybrid_0.1
Hybrid pipeline software specifications: autoai-kb_r24.1-py3.11

Tags
Add tags to make assets easier to find.

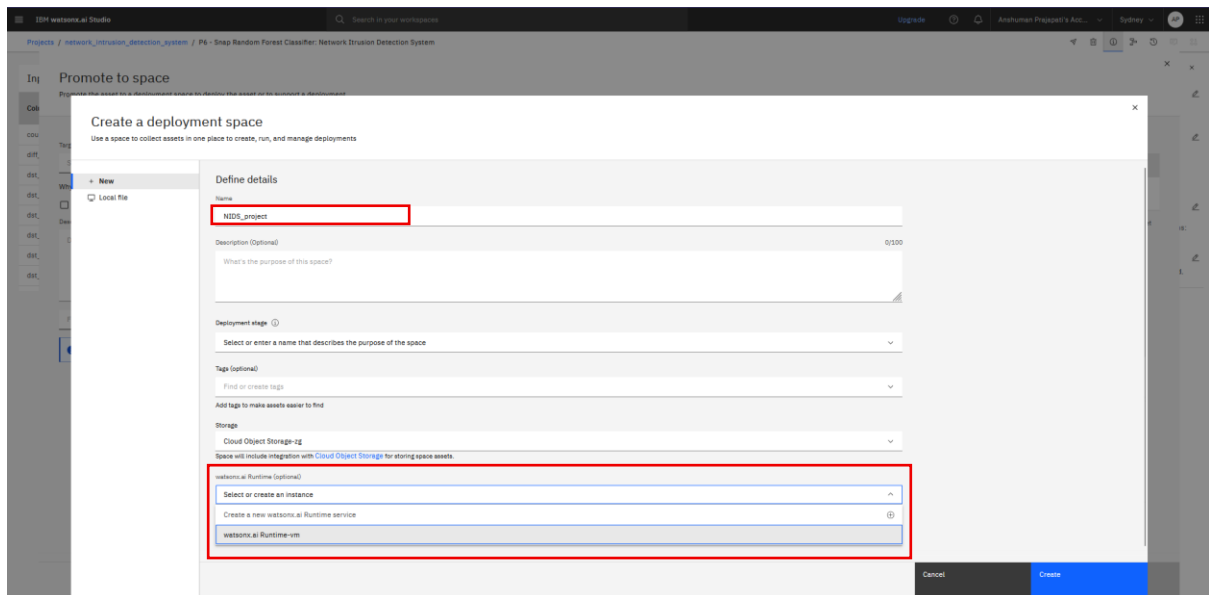
Last modified
1 minute ago by Service

Created on
Aug 3, 2025 by Anshuman Prajapati

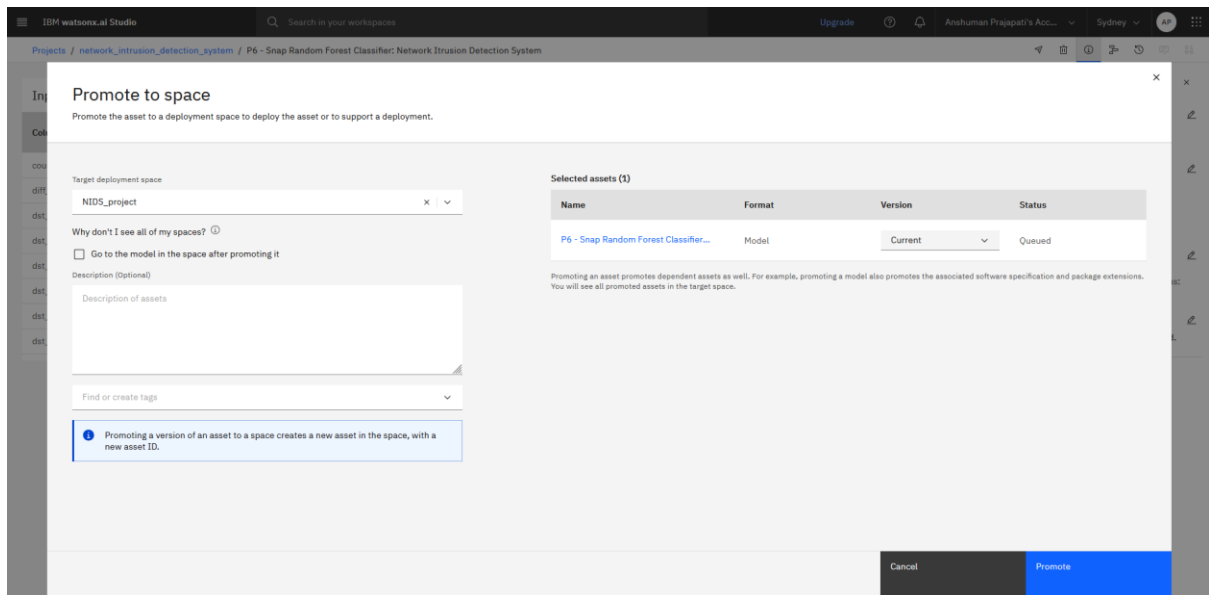
30. Now on the top towards right press on promote to space.



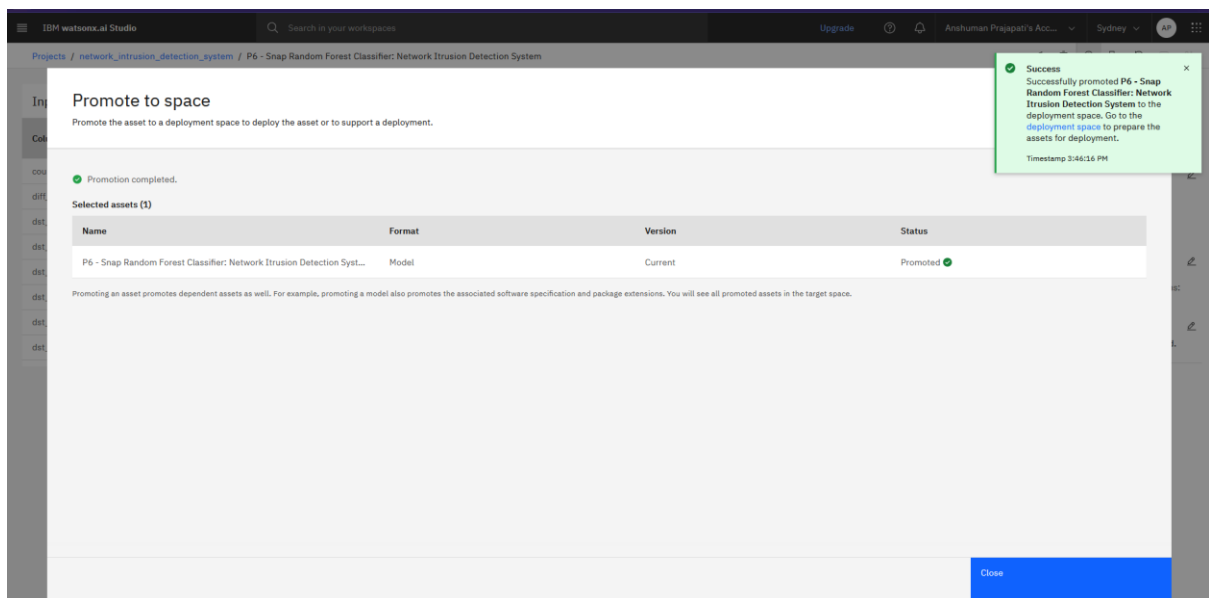
31. Select create new deployment .



32. Name the deployment space and make sure to select instance of watsonx.ai Runtime and press create.



33. Now press on promote.



34. Now after it got promoted to space press on deployment space ,the pop up on top right corner.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Acc...

Sydney

Deployment spaces /

NIDS_project

Overview Assets Deployments Jobs Manage

Find assets

Import assets

New asset

1 asset

All assets

Asset types

Models

Name	Last modified
P6 - Snap Random Forest Classifier: Network Intrusion Detection System Machine learning model from AutoAI	1 minute ago Service

Items per page: 20 1-5 of 1 items

1 of 1 pages

35. Now click on the asset.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Acc...

Sydney

Deployment spaces / NIDS_project / P6 - Snap Random Forest Classifier: Network Intrusion Detection System

Deployments Model details

Search

New deployment

Name	Type	Status	Tags	Last modified
<p>This asset doesn't have any deployments yet</p> <p>Use the New Deployment button to create a deployment for this asset.</p>				

Items per page: 20 0-0 of 0 items

1 of 1 pages

About this asset

Name

P6 - Snap Random Forest Classifier: Network Intrusion Detection System

Description

No description provided.

Asset Details

Type: wml-hybrid_0.1

Model ID: ab45f621-f1a9-44...

Software specification: hybrid_0.1

Hybrid pipeline software specifications: autoai-kb_r24.1-py3.11

Tags

Add tags to make assets easier to find.

Source asset details

Last modified

3 minutes ago by Service

Created on

Aug 3, 2025 by Anshuman Prajapati

36. Now click on new **New Deployment**

Create a deployment

Define details

Associated asset
P6 - Snap Random Forest Classifier: Network Intrusion Detection System

Deployment type

Online
Run the model on data in real-time, as data is received by a web service.

Batch
Run the model against data as a batch process.

Name
NIDS_final

Serving name
Deployment serving name
Enter a short name to be used as the serving name for the deployment. The name must be unique to be valid.

Description
Deployment description

Tags

Cancel Create

37. Now select online and give this final deployment a name and press create.

Deployments

Search

New deployment

Name	Type	Status	Tags	Last modified
(y) NIDS_final	Online	Deployed		29 seconds ago Anshuman Prajapati (You)

Items per page: 20 1-1 of 1 items 1 of 1 pages

About this asset

Name
P6 - Snap Random Forest Classifier: Network Intrusion Detection System

Description
No description provided.

Asset Details
Type: wml-hybrid_0.1
Model ID: ab45621-f1a9-44...
Software specification: hybrid_0.1
Hybrid pipeline software specifications: autoai-kb_r24.1-py3.11

Tags
Add tags to make assets easier to find.

Source asset details

Last modified
5 minutes ago by Service

Created on
Aug 3, 2025 by Anshuman Prajapati

38. After some time the deployment will be completed, then press on the deployment name.

The screenshot shows the IBM watsonx.ai Studio interface. The main content area displays the 'API reference' for a deployment named 'NIDS_final'. It includes sections for 'Endpoints for scoring' with private and public endpoints, and 'Code snippets' for various languages. A red box highlights the 'Test' button in the API reference section. The right sidebar provides details about the deployment, including its name, description, and associated assets.

39. This will give the API reference page where API's are present to implement this model with other projects. Now press on test to test the model.

The screenshot shows the 'Enter input data' page in IBM watsonx.ai Studio. It includes a text input field with the value 'JSON', a 'Test' button, and a large table for entering data manually. The table has columns for various input features and a 'Predict' button at the bottom right.

	duration (double)	protocol_type (other)	service (other)	flag (other)	src_bytes (double)	dst_bytes (double)	land (double)	wrong_fragment (double)	urgent (double)	hot (double)	num_failed_logins (double)	logged_in
1	Start typing or drag and drop a CSV file...											
2												
3												
4												
5												
6												
7												
8												
9												
10												

40. We have many options to give input for test ,I choose local file as I have one.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Anshuman Prajapati's Acc...

Sydney

Deployment spaces / NIDS_project / P6 - Snap Random Forest Classifier: Network Intrusion Detection System /

NIDS_final Deployed Online

API reference **Test**

Enter input data

Text JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

[Download CSV template](#) [Browse local files](#) [Search in space](#) [Clear all](#) X

	duration (double)	protocol_type (other)	service (other)	flag (other)	src_bytes (double)	dst_bytes (double)	land (double)	wrong_fragment (double)	urgent (double)	hot (double)	num_failed_logins (double)	logged_in
1	0	tcp	private	REJ	0	0	0	0	0	0	0	0
2	0	tcp	private	REJ	0	0	0	0	0	0	0	0
3	2	tcp	ftp_data	SF	12983	0	0	0	0	0	0	0
4	0	icmp	eco_j	SF	20	0	0	0	0	0	0	0
5	1	tcp	telnet	RSTO	0	15	0	0	0	0	0	0
6	0	tcp	http	SF	267	14515	0	0	0	0	0	1
7	0	tcp	smtp	SF	1022	387	0	0	0	0	0	1
8	0	tcp	telnet	SF	129	174	0	0	0	0	1	0
9	0	tcp	http	SF	327	467	0	0	0	0	0	1
10	0	tcp	ftp	SF	26	157	0	0	0	0	1	0
11	0	tcp	telnet	SF	0	0	0	0	0	0	0	0
22,544 rows, 41 columns												

Predict

41. After the file is loaded press predict. It will take some time to generate the result.

Prediction results

Prediction type

Binary classification

Prediction percentage

22,544 records

■ anomaly ■ normal

Confidence level distribution

Display format for prediction results

☒ Table view ☐ JSON view ☐ Show input data ⓘ

	Prediction	Confidence
1	anomaly	100%
2	anomaly	100%
3	normal	100%
4	anomaly	100%
5	normal	100%
6	normal	100%
7	normal	100%
8	normal	100%
9	normal	100%
10	anomaly	100%
11	anomaly	100%
12	normal	100%
13	anomaly	100%
14	anomaly	100%
15	normal	100%
16	normal	100%
17	normal	100%

Download JSON file

42. This will generate a final result with up to 100% accuracy.

For upgrade of different types of attack detection just the data(.csv file) on which model is trained is to be changed for new data(enhanced.csv file) provided above and also you can refer this [GitHub](#) repo (the enhanced files).