CAPSTONE PROJECT

NETWORK INTRUSION DETECTION

Presented By:

Student Name :- Divyanshu Roy
College Name :- SANDIP UNIVERSITY
Department :- BTech(CSE)
AICTE Student ID: STU6857a2a6d554c1750573734
Internship ID: INTERNSHIP_175040918768551fe3c08b0
Technology: IBM Watsonx.ai Studio

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Network Intrusion Detection The Challenge:

Create a robust network intrusion detection system (NIDS) using machine learning. The system should be capable of analyzing 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.

Kaggle dataset link – https://www.kaggle.com/datasets/sampadab17/network intrusion-detection

PROPOSED SOLUTION

The proposed solution is an intelligent system built on the IBM Cloud platform that leverages machine learning to automate threat detection.

- **Data Source:** Utilizes the well-known NSL-KDD dataset from Kaggle, which contains a wide variety of network intrusions.
- **Automated Model Building:** Employs the **AutoAI** feature within **IBM Watsonx.ai** to automatically preprocess the data, select the best classification algorithm, and optimize its performance.
- **Prediction Goal:** The model will be trained to predict the **'class'** of network activity (e.g., 'normal', 'dos', 'probe', etc.).
- **Deployment:** The final, most accurate model will be deployed as a live web service (API), capable of making real-time predictions on new network data.

SYSTEM APPROACH

This project was developed using a suite of powerful cloud-based AI tools:

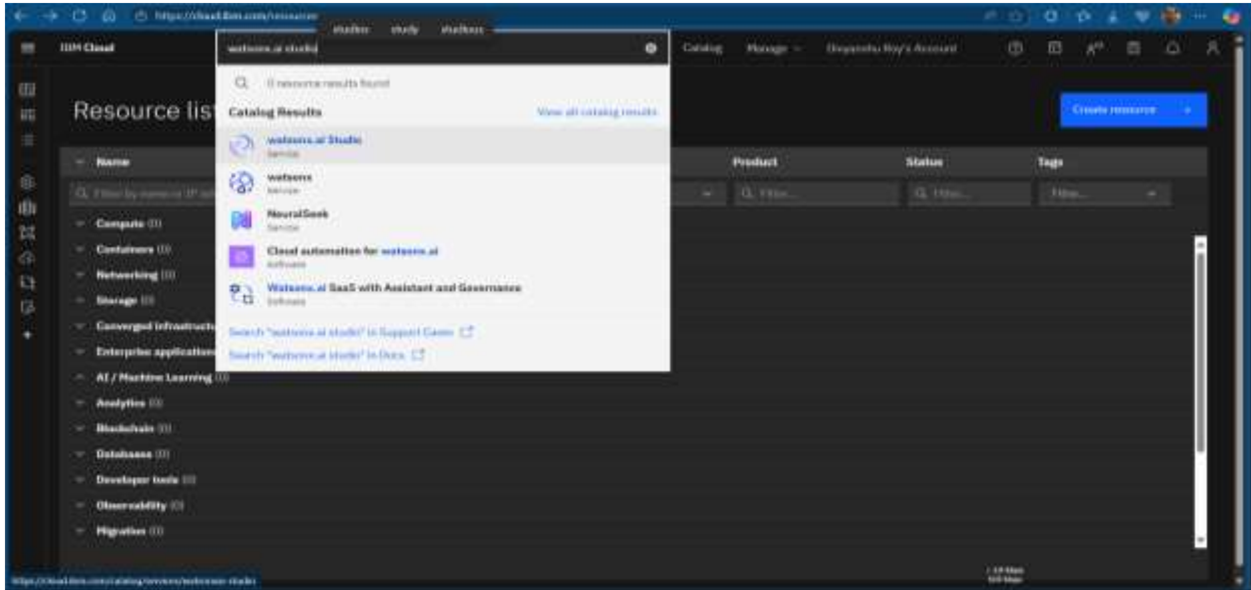
- **Cloud Platform:** IBM Cloud
- **AI/ML Studio:** IBM Watsonx.ai
- **Core Engine:** AutoAI Experiment
- **Model Deployment:** Watson Machine Learning Service
- **Dataset:** NSL-KDD Network Intrusion Dataset (from Kaggle)
    - Kaggle dataset link – https://www.kaggle.com/datasets/sampadab17/network intrusion-detection

ALGORITHM & DEPLOYMENT

The project was executed following a precise, step-by-step workflow within the IBM Cloud environment:

1. Logged into the **IBM Cloud** platform.
2. Cleared the resource list to ensure a clean workspace.
3. Created a **New Project** in Watsonx.ai, configuring the necessary runtime and storage services.
4. Navigated to the "Build machine learning model automatically" section.
5. Configured the **AutoAI Experiment** with a name and description.
6. Uploaded the **Train_data.csv** as the data source.
7. **Ran the experiment**, which automatically trained and evaluated multiple models.
8. Selected and **saved the pipeline** with the highest accuracy from the results.
9. **Promoted the model** to a deployment space and deployed it as a live service.
10. **Tested** the deployed model to ensure it was making predictions correctly.

SCREENSHOTS OF WORKFLOW



Resource list

watsonx.ai studio

0 resource results found

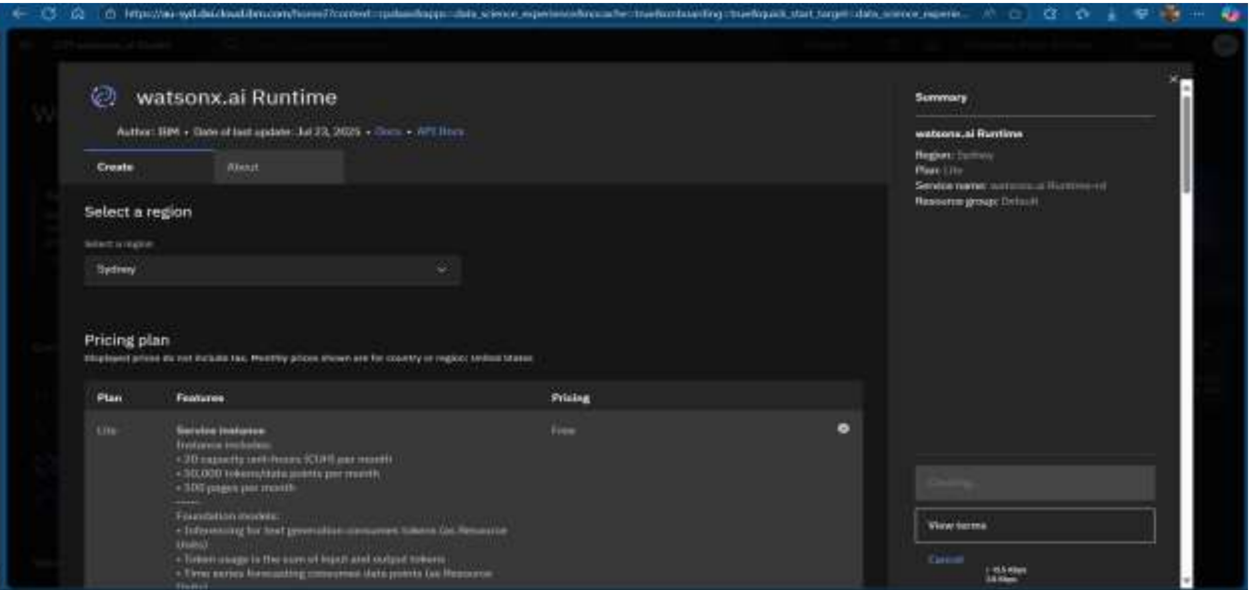
Catalog Results — View all catalog results

- watsonx.ai Studio — Service
- watsonx — Service
- NeuralSeek — Service
- Cloud automation for watsonx.ai — Software
- Watsonx.ai SaaS with Assistant and Governance — Software

Search "watsonx.ai studio" in Support Cases

Search "watsonx.ai studio" in Docs

Compute (0)
Containers (0)
Networking (0)
Storage (0)
Converged infrastructure
Enterprise applications
AI / Machine Learning (1)
Analytics (0)
Blockchain (0)
Databases (0)
Developer tools (0)
Observability (0)
Migration (0)



watsonx.ai Runtime

Author: IBM • Date of last update: Jul 23, 2025 • Docs • API Docs

Select a region

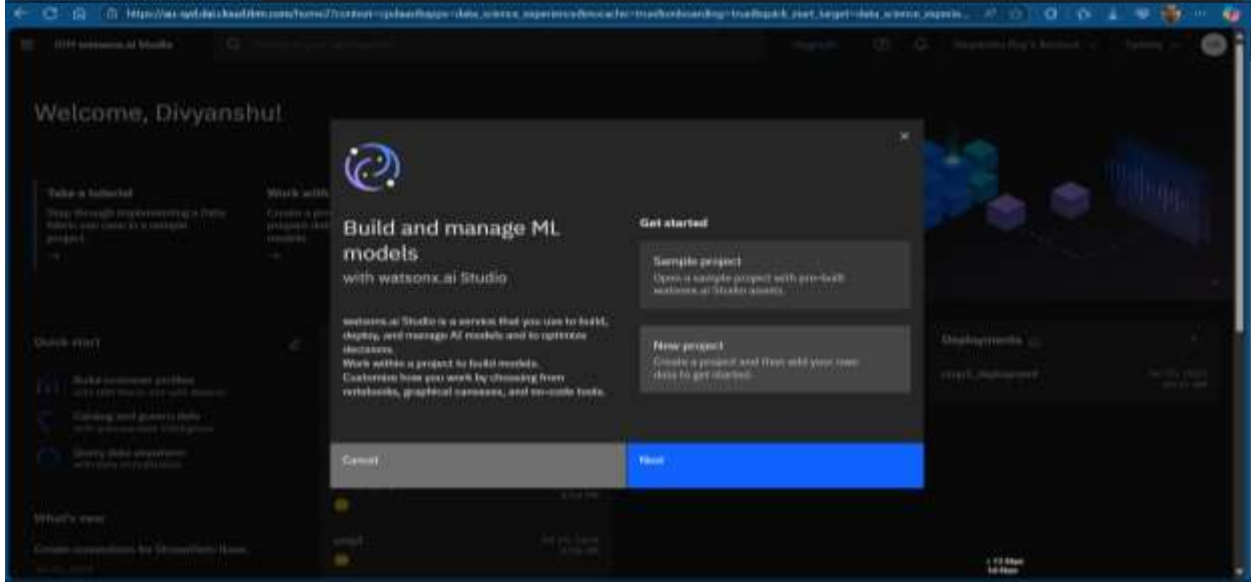
Select a region: Sydney

Pricing plan

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

Plan	Features	Pricing
Lite	Service instance Instance includes: • 20 capacity unit-hours (CUH) per month • 50,000 tokens/data points per month • 500 pages per month Foundation models: • Inferencing for text generation consumes tokens (as Resource Units) • Token usage is the sum of input and output tokens • Time series forecasting consumes data points (as Resource Units)	Free

watsonx.ai Runtime
Region: Sydney
Plan: Lite
Service name: watsonx.ai Runtime-rd
Resource group: Default



Welcome, Divyanshu!

Build and manage ML models

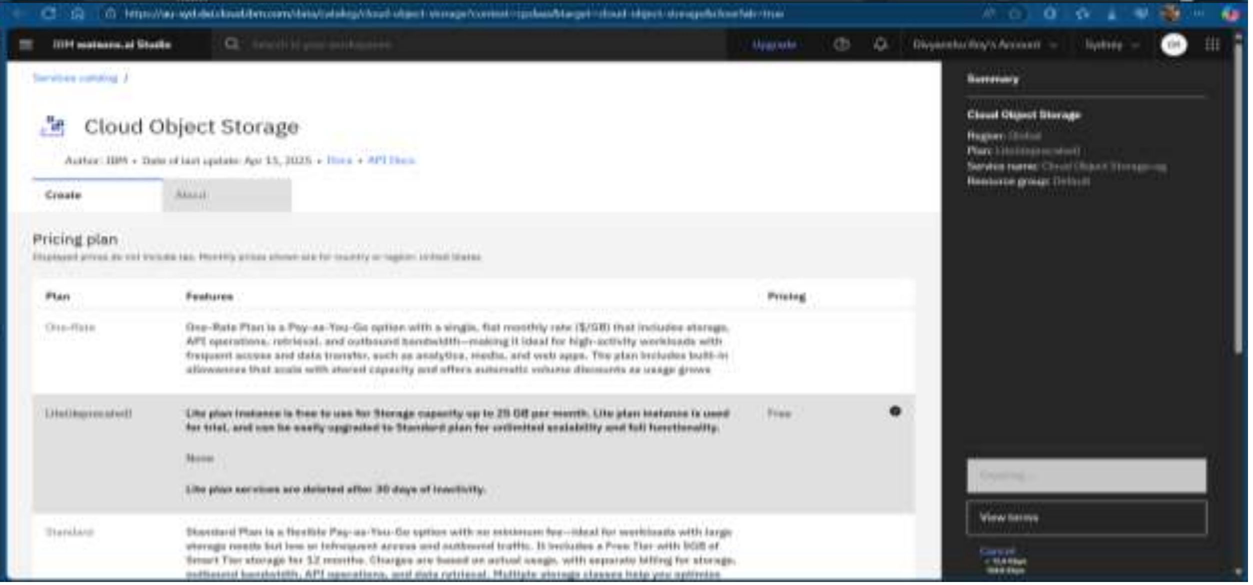
with watsonx.ai Studio

watsonx.ai Studio is a service that you use to build, deploy, and manage AI models and to optimize decisions.
Work within a project to build models.
Customize how you work by choosing from notebooks, graphical canvases, and no-code tools.

Get started

Sample project — Open a sample project with pre-built watsonx.ai Studio assets.

New project — Create a project and then add your own data to get started.



Cloud Object Storage

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

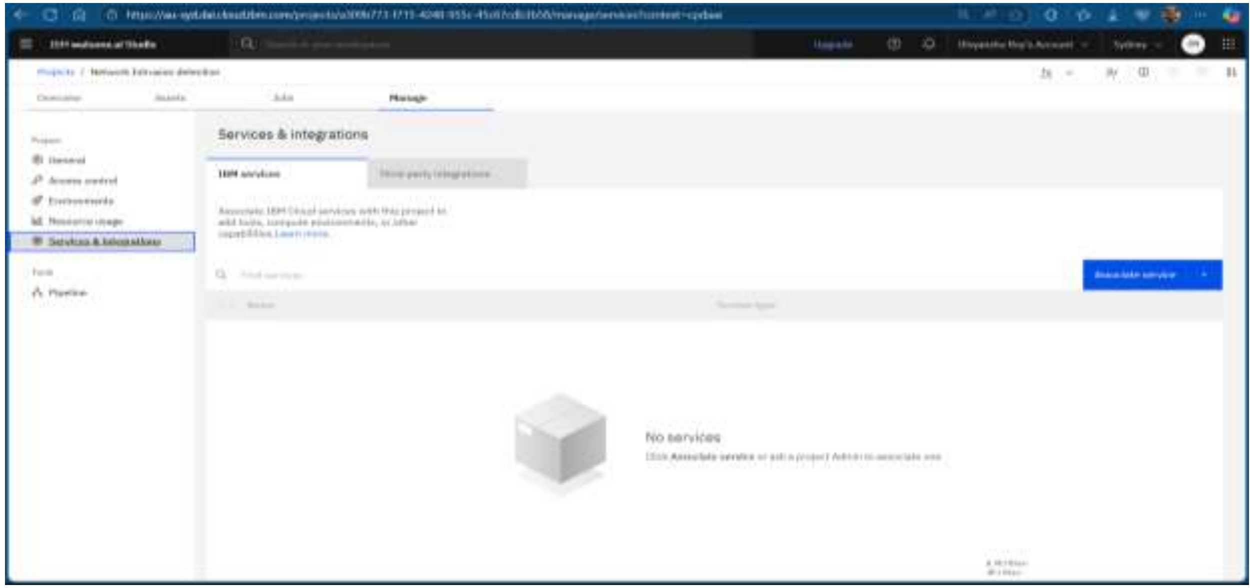
Pricing plan

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)	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. None Lite plan services are deleted after 30 days of inactivity.	Free
Standard	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	

Cloud Object Storage
Region: Global
Plan: Lite(deprecated)
Service name: Cloud Object Storage-eg
Resource group: Default

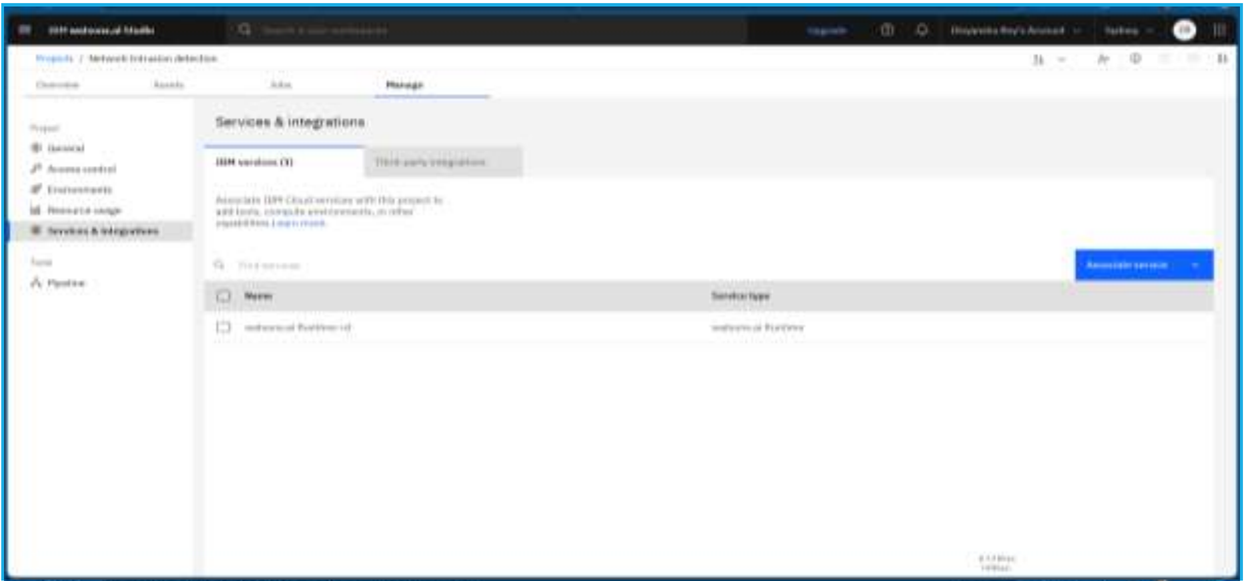
SCREENSHOTS OF WORKFLOW



Services & integrations

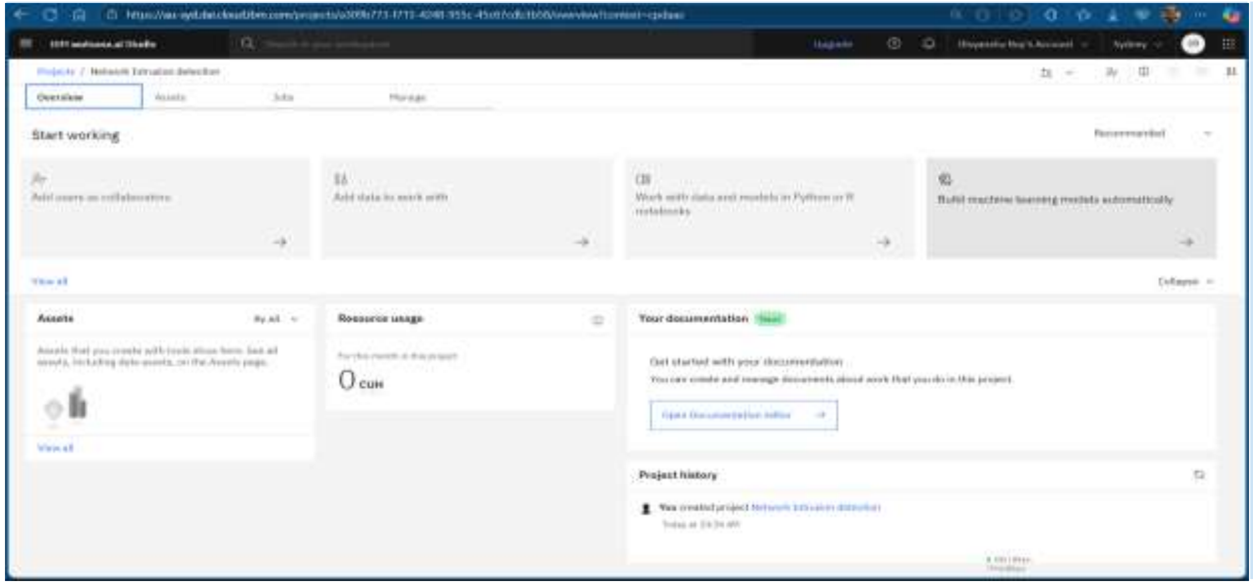
Associate IBM Cloud services with this project to add tools, compute environments, or other capabilities. Learn more.

No services



Services & integrations

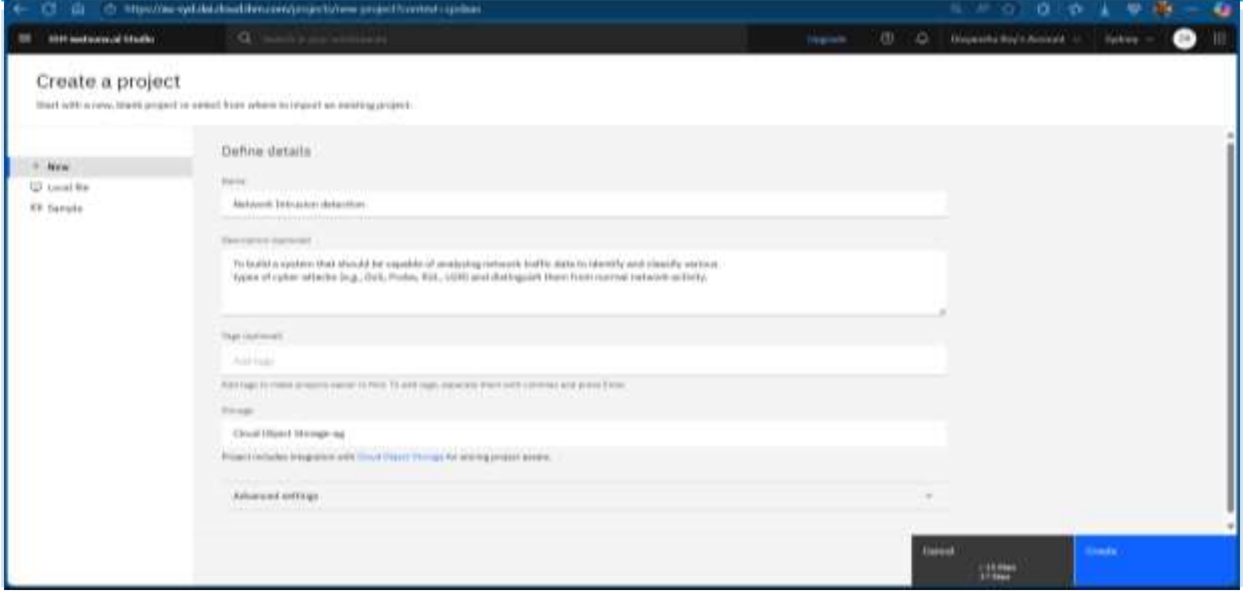
Associate IBM Cloud services with this project to add tools, compute environments, or other capabilities. Learn more.



Start working

Get started with your documentation

You created project Network Intrusion detection



Create a project

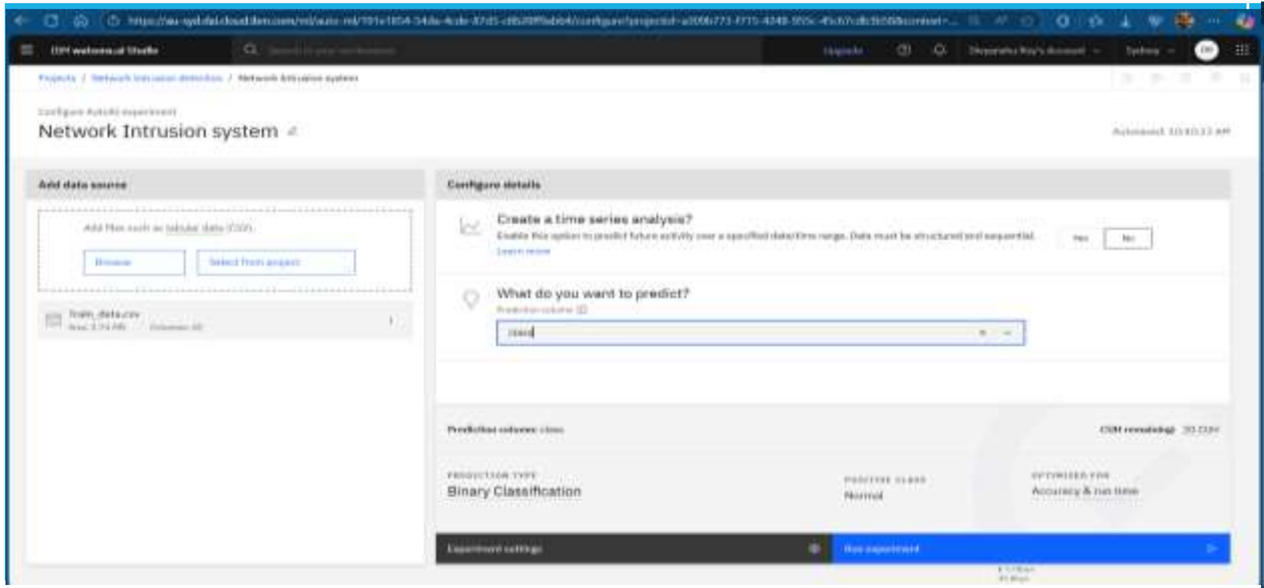
Define details

Name: Network Intrusion detection

Description (optional): To build a system that should be capable of analyzing 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.

Storage: Cloud Object Storage

SCREENSHOTS OF WORKFLOW



Network Intrusion system

Add data source

Configure details

Create a time series analysis?

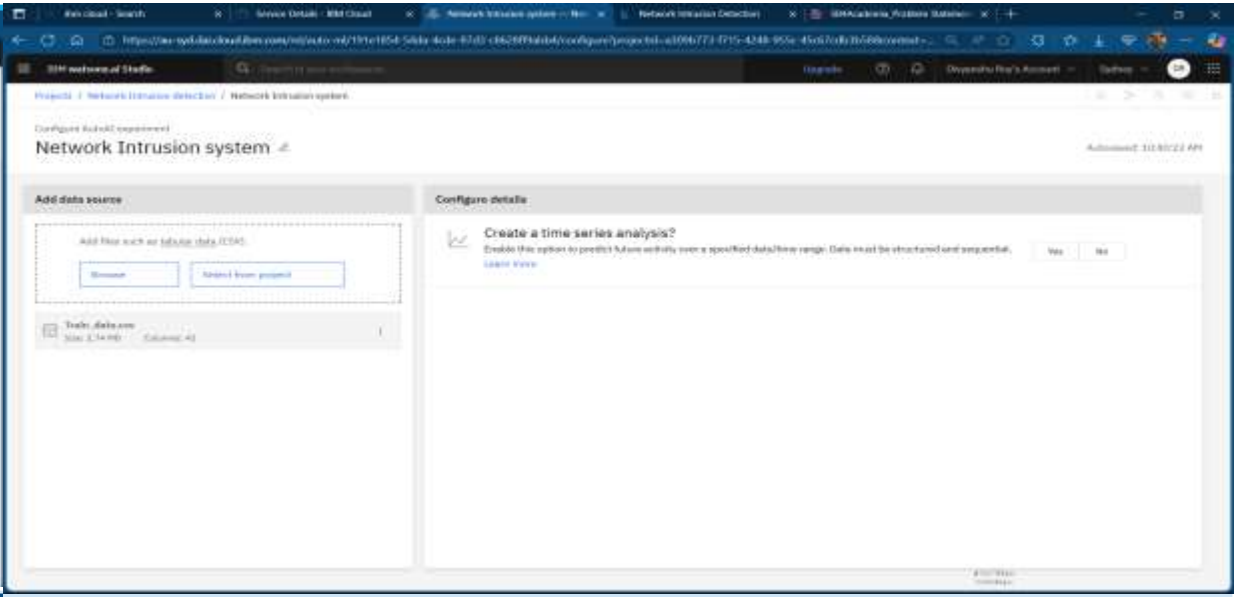
What do you want to predict?

Prediction column: class

PREDICTION TYPE: Binary Classification

POSITIVE CLASS: Normal

OPTIMIZED FOR: Accuracy & run time

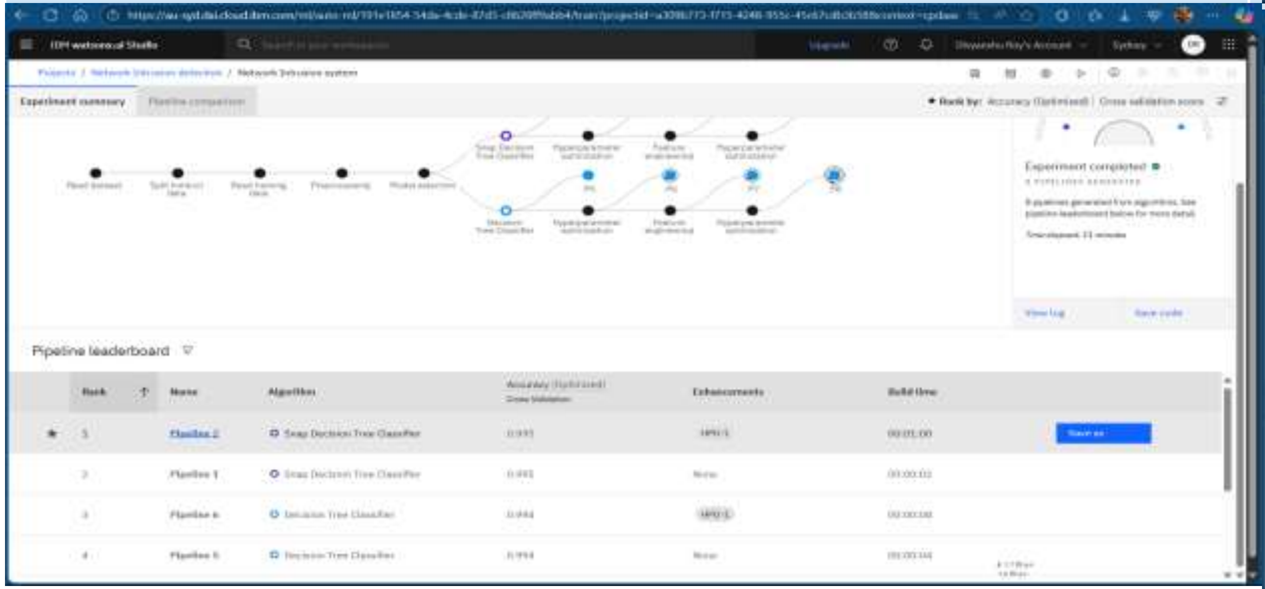


Network Intrusion system

Add data source

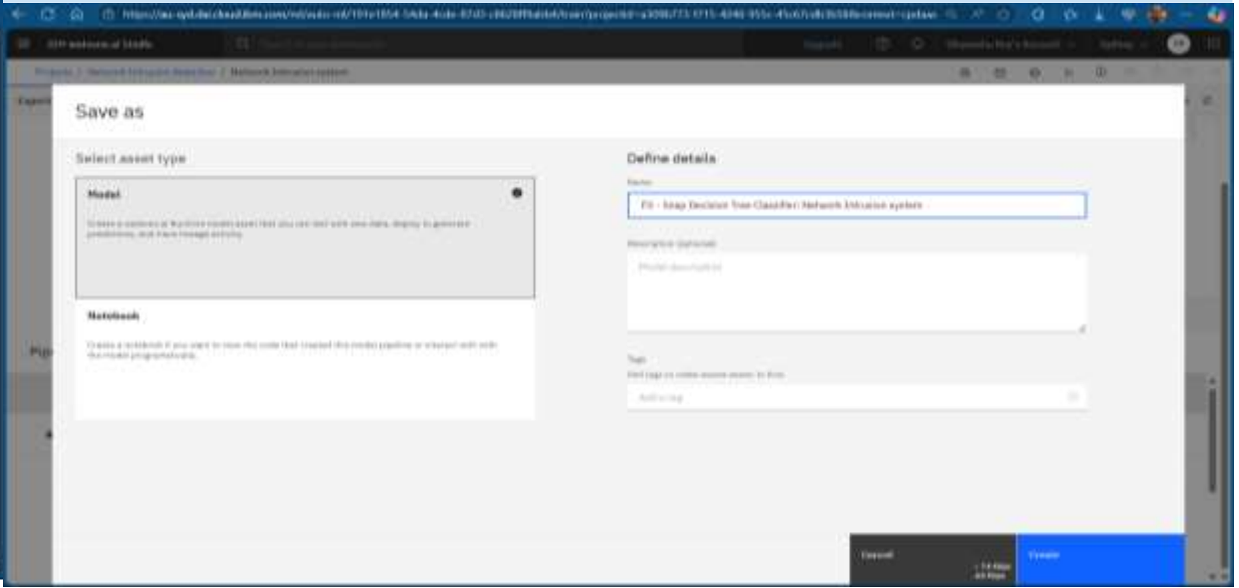
Configure details

Create a time series analysis?



Pipeline leaderboard

Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
1	Pipeline 2	Snap Decision Tree Classifier	[illegible]	HPO-1	00:01:00
2	Pipeline 1	Snap Decision Tree Classifier	[illegible]	None	[illegible]
3	Pipeline 4	Decision Tree Classifier	[illegible]	HPO-1	[illegible]
4	Pipeline 5	Decision Tree Classifier	[illegible]	None	[illegible]

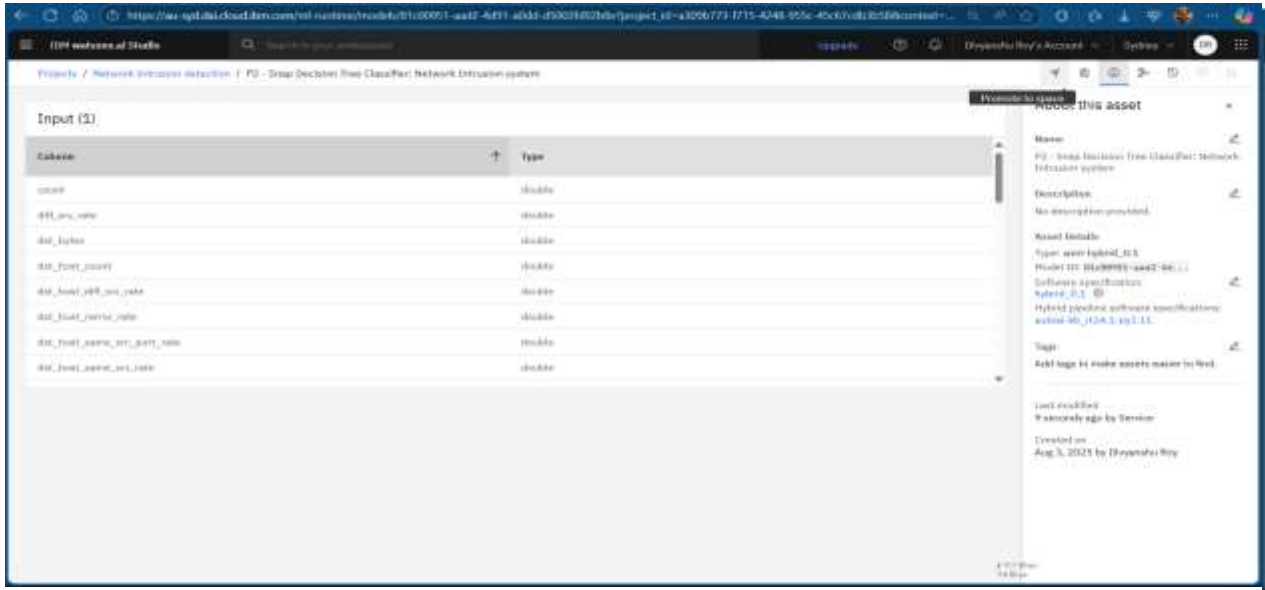


Save as

Select asset type

Define details

SCREENSHOTS OF WORKFLOW



IBM watsonx.ai Studio

Projects / Network intrusion detection / P2 - Snap Decision Tree Classifier: Network Intrusion system

Promote to space

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_rerror_rate	double
dst_host_same_src_port_rate	double
dst_host_same_srv_rate	double

About this asset

Name: P2 - Snap Decision Tree Classifier: Network Intrusion system

Description: No description provided.

Asset Details

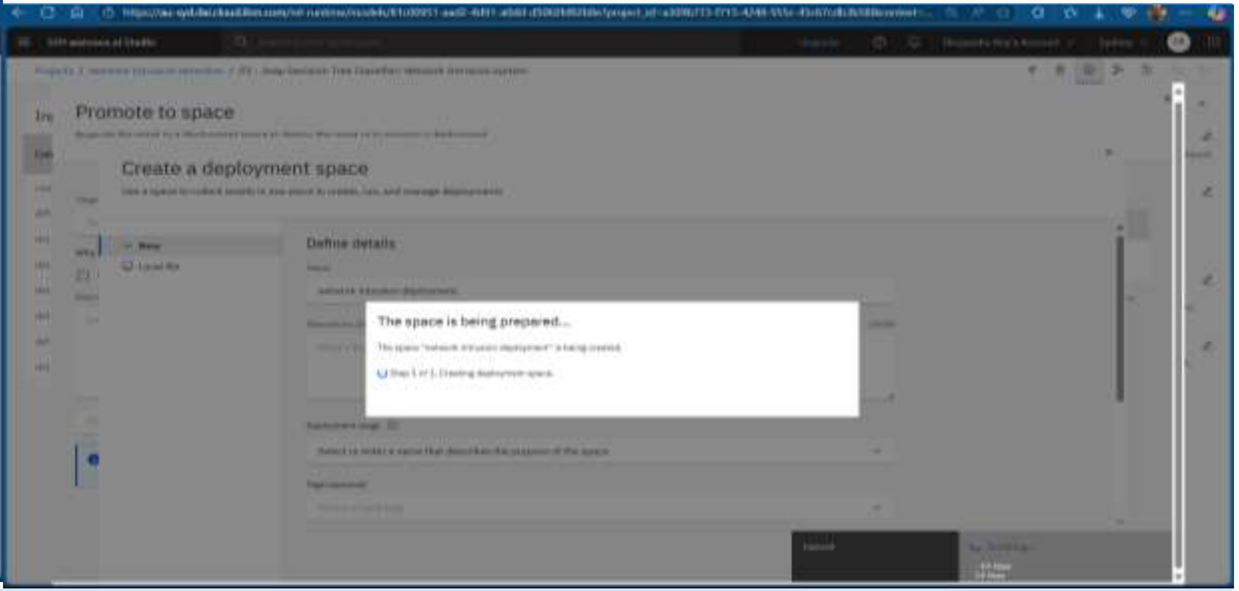
Type: wml-hybrid_0.1

Software specification: hybrid_0.1

Tags: Add tags to make assets easier to find.

Last modified: 9 seconds ago by Service

Created on: Aug 3, 2025 by Divyanshu Roy



Promote to space

Create a deployment space

Use a space to collect assets in one place to create, run, and manage deployments.

New

Local file

Define details

Name: network intrusion deployment

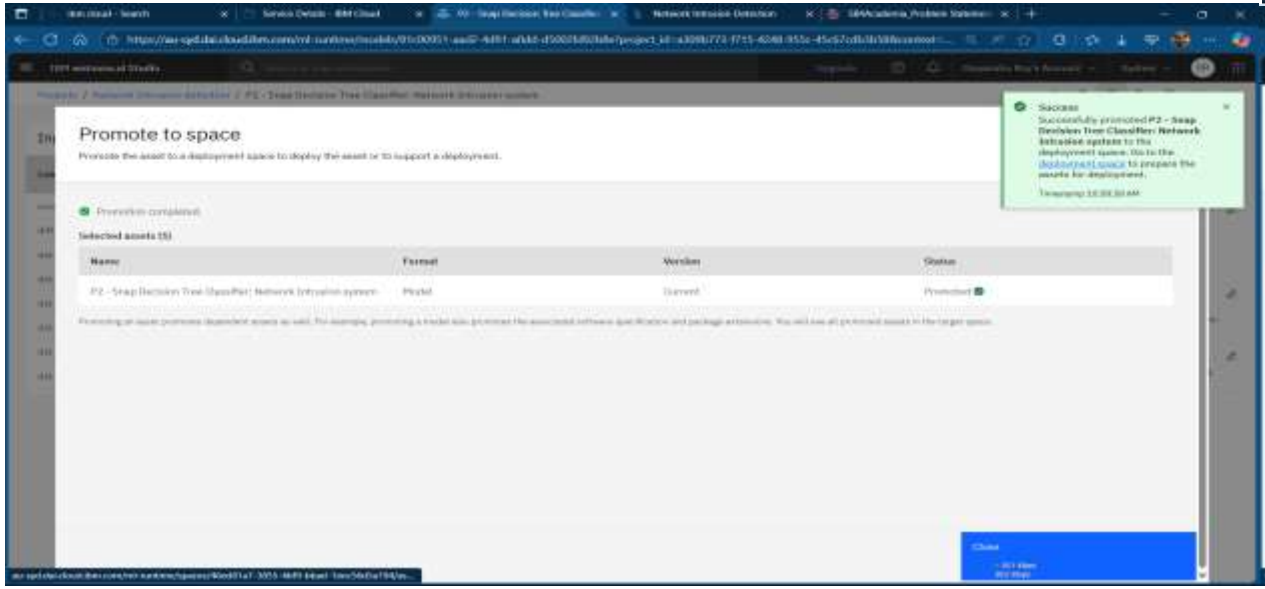
The space is being prepared...

The space "network intrusion deployment" is being created.

Step 1 of 1: Creating deployment space

Deployment stage: Select or enter a name that describes the purpose of the space

Cancel



Promote to space

Promote the asset to a deployment space to deploy the asset or to support a deployment.

Promotion completed

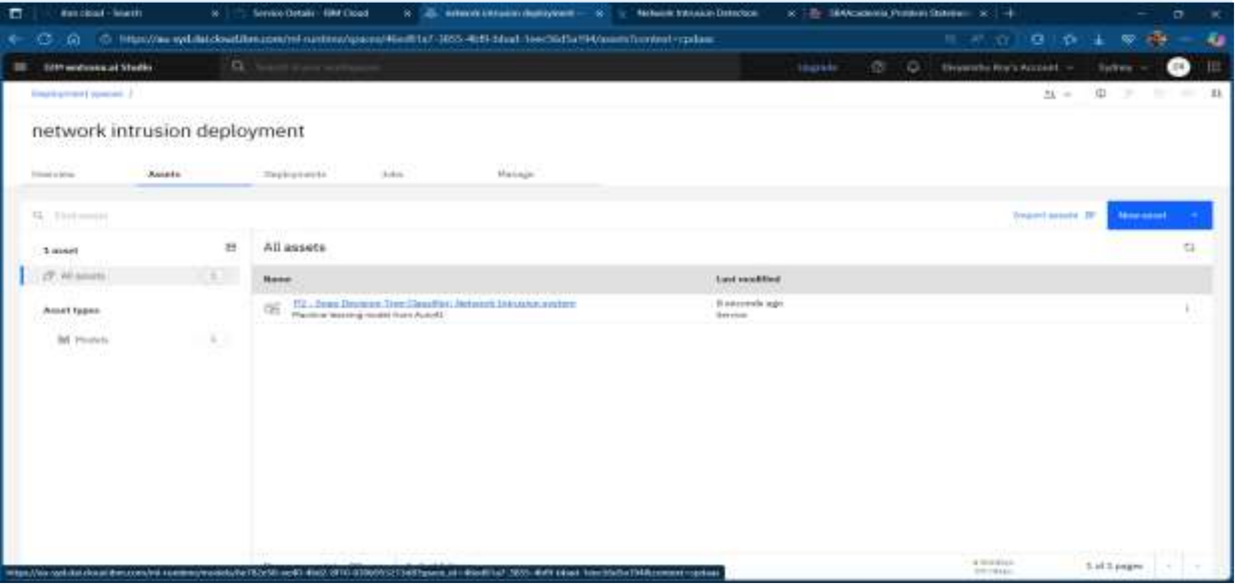
Selected assets (1)

Name	Format	Version	Status
P2 - Snap Decision Tree Classifier: Network Intrusion system	Model	Current	Promoted

Success

Successfully promoted **P2 - Snap Decision Tree Classifier: Network Intrusion system** to the deployment space. Go to the deployment space to prepare the assets for deployment.

Close



Deployment spaces /

network intrusion deployment

Overview | Assets | Deployments | Jobs | Manage

Import assets | New asset

1 asset

All assets 1

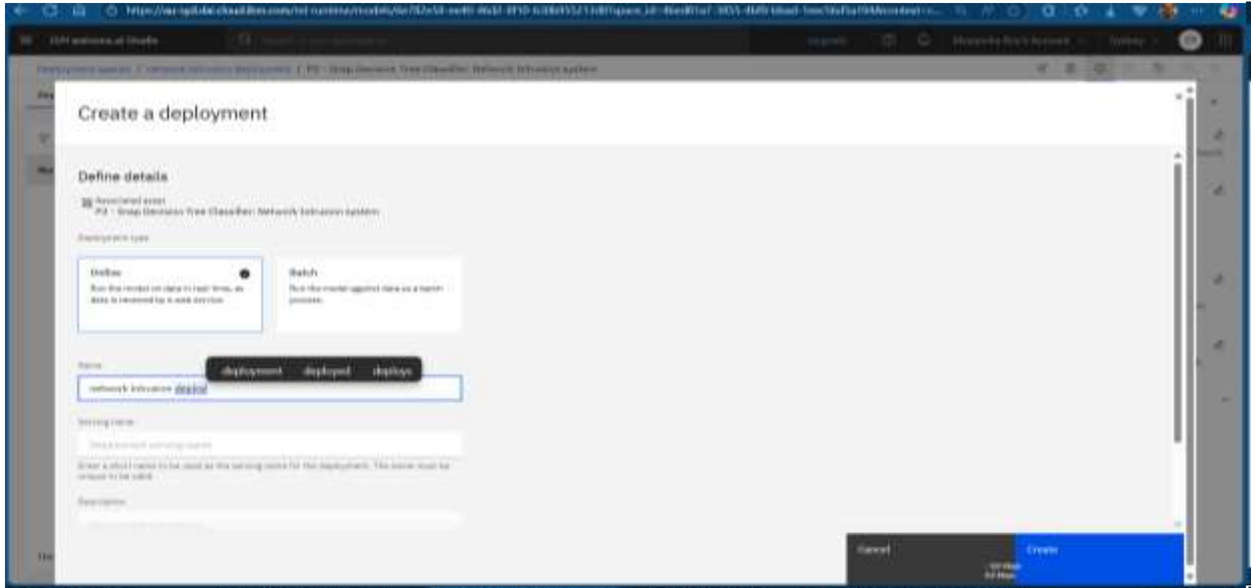
Asset types

Models 1

All assets

Name	Last modified
P2 - Snap Decision Tree Classifier: Network Intrusion system Machine learning model from AutoAI	8 seconds ago Service

SCREENSHOTS OF WORKFLOW



IBM watsonx.ai Studio

Deployment spaces / network intrusion deployment / P2 - Snap Decision Tree Classifier: Network Intrusion system

Create a deployment

Define details

Associated asset
P2 - Snap Decision Tree Classifier: Network Intrusion system

Deployment type

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

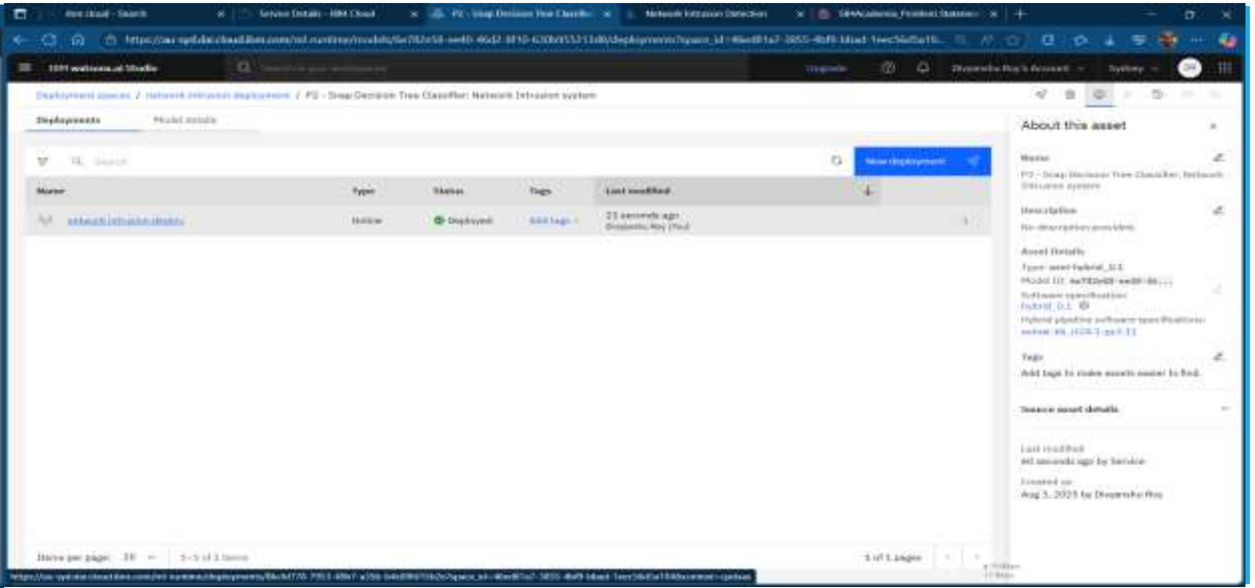
Batch
Run the model against data as a batch process.

Name
network intrusion deploy

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



IBM watsonx.ai Studio

Deployment spaces / Network intrusion deployment / P2 - Snap Decision Tree Classifier: Network Intrusion system

Name	Type	Status	Tags	Last modified
network intrusion deploy	Online	Deployed	Add tags +	23 seconds ago Divyanshu Roy (You)

Items per page: 10 | 1–1 of 1 items | 1 of 1 pages

About this asset

Name
P2 - Snap Decision Tree Classifier: Network Intrusion system

Description
No description provided.

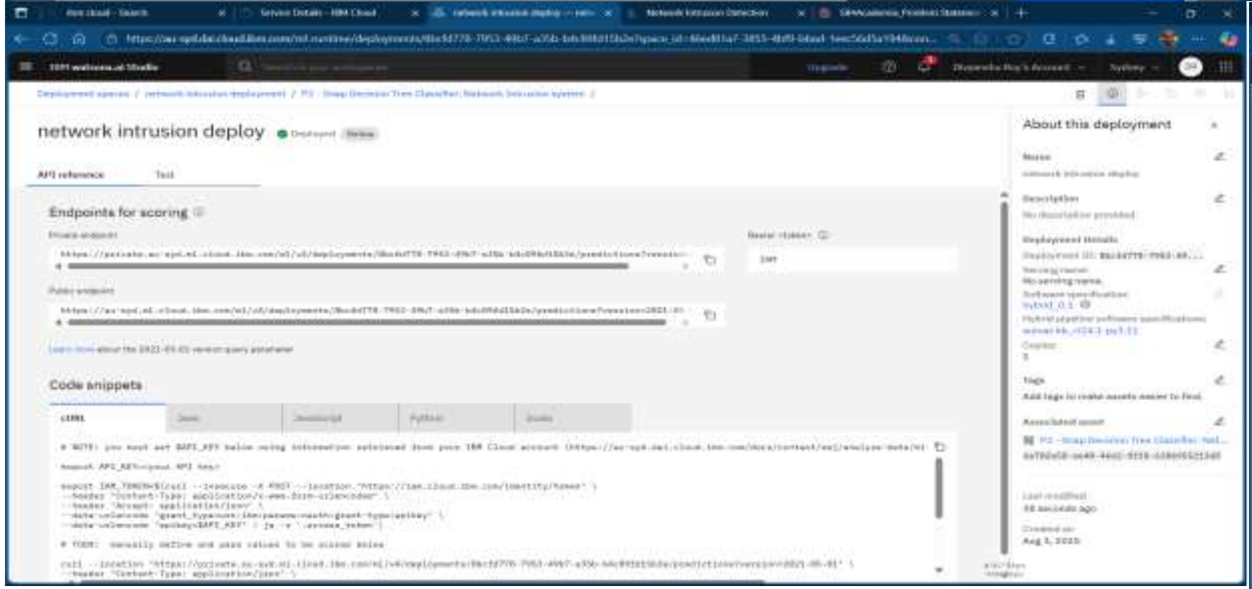
Asset Details
Type: wml-hybrid_0.1
Model ID: 6e782e58-ee40-46...
Software specification:
hybrid_0.1
Hybrid pipeline software specifications:
autoai-kb_rt24.1-py3.11

Tags
Add tags to make assets easier to find.

Source asset details

Last modified
44 seconds ago by Service

Created on
Aug 3, 2025 by Divyanshu Roy



IBM watsonx.ai Studio

Deployment spaces / network intrusion deployment / P2 - Snap Decision Tree Classifier: Network Intrusion system

network intrusion deploy ● Deployed Online

Endpoints for scoring

Private endpoint
https://private.au-syd.ml.cloud.ibm.com/ml/v4/deployments/8bcfd778-7953-49b7-a35b-b4c898d15b2e/predictions?versio...

Bearer token
IAM

Public endpoint
https://au-syd.ml.cloud.ibm.com/ml/v4/deployments/8bcfd778-7953-49b7-a35b-b4c898d15b2e/predictions?version=2021-05-...

Learn more about the 2021-05-01 version query parameter

Code snippets

```
# NOTE: you must set $API_KEY below using information retrieved from your IBM Cloud account (https://au-syd.dai.cloud.ibm.com/docs/content/wsj/analyze-data/ml
export API_KEY=<your API key>

export IAM_TOKEN=$(curl --insecure -X POST --location "https://iam.cloud.ibm.com/identity/token" \
--header "Content-Type: application/x-www-form-urlencoded" \
--header "Accept: application/json" \
--data-urlencode "grant_type=urn:ibm:params:oauth:grant-type:apikey" \
--data-urlencode "apikey=$API_KEY" | jq -r '.access_token')

# TODO:  manually define and pass values to be scored below

curl --location "https://private.au-syd.ml.cloud.ibm.com/ml/v4/deployments/8bcfd778-7953-49b7-a35b-b4c898d15b2e/predictions?version=2021-05-01" \
--header "Content-Type: application/json" \
```

About this deployment

Name
network intrusion deploy

Description
No description provided.

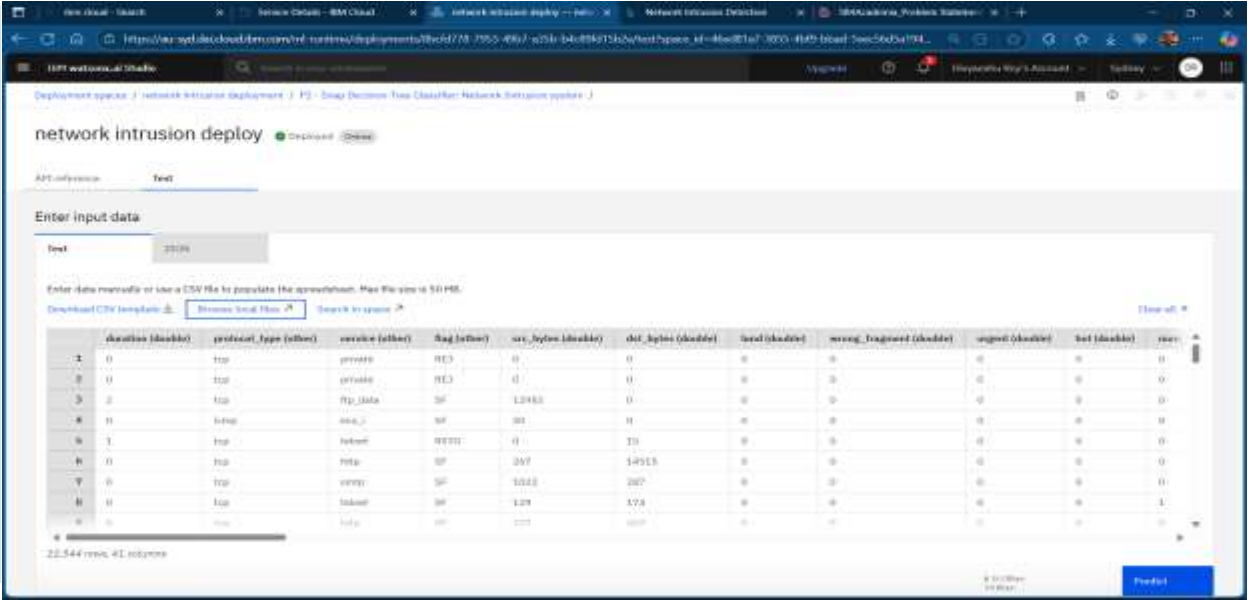
Deployment Details
Deployment ID: 8bcfd778-7953-49...
Serving name:
No serving name.
Software specification:
hybrid_0.1
Hybrid pipeline software specifications:
autoai-kb_rt24.1-py3.11
Copies:
1

Tags
Add tags to make assets easier to find.

Associated asset
P2 - Snap Decision Tree Classifier: Net...
6e782e58-ee40-46d2-8f10-6308f5521348

Last modified
38 seconds ago

Created on
Aug 3, 2025



IBM watsonx.ai Studio

Deployment spaces / network intrusion deployment / P2 - Snap Decision Tree Classifier: Network Intrusion system

network intrusion deploy ● Deployed Online

Enter input data

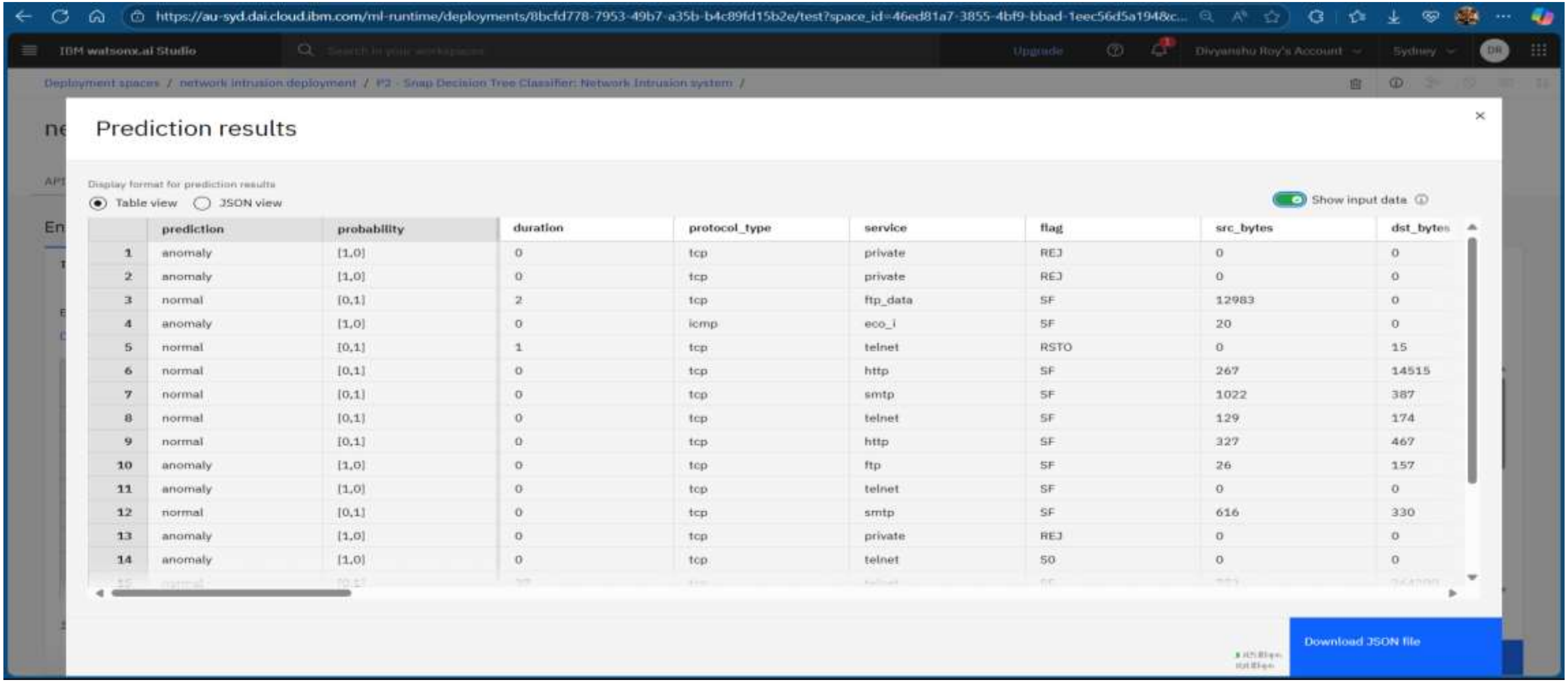
Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

	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_
1	0	tcp	private	REJ	0	0	0	0	0	0	0
2	0	tcp	private	REJ	0	0	0	0	0	0	0
3	2	tcp	ftp_data	SF	12983	0	0	0	0	0	0
4	0	icmp	eco_i	SF	20	0	0	0	0	0	0
5	1	tcp	telnet	RSTO	0	15	0	0	0	0	0
6	0	tcp	http	SF	267	14515	0	0	0	0	0
7	0	tcp	smtp	SF	1022	387	0	0	0	0	0
8	0	tcp	telnet	SF	129	174	0	0	0	0	1

22,544 rows, 41 columns

RESULT

The AutoAI experiment successfully generated multiple pipelines, with the top-performing model (**Pipeline 2**) achieving an accuracy of **99.5%**. The model was then deployed and tested, correctly identifying network traffic as 'normal' or 'anomaly'.



https://au-syd.dai.cloud.ibm.com/ml-runtime/deployments/8bcfd778-7953-49b7-a35b-b4c89fd15b2e/test?space_id=46ed81a7-3855-4bf9-bbad-1eec56d5a194&c...
IBM watsonx.ai Studio
Upgrade
Divyanshu Roy's Account
Sydney
DR
Deployment spaces / network intrusion deployment / P2 - Snap Decision Tree Classifier: Network Intrusion system /

Prediction results

Display format for prediction results

● Table view ○ JSON view

Show input data

	prediction	probability	duration	protocol_type	service	flag	src_bytes	dst_bytes
1	anomaly	[1,0]	0	tcp	private	REJ	0	0
2	anomaly	[1,0]	0	tcp	private	REJ	0	0
3	normal	[0,1]	2	tcp	ftp_data	SF	12983	0
4	anomaly	[1,0]	0	icmp	eco_i	SF	20	0
5	normal	[0,1]	1	tcp	telnet	RSTO	0	15
6	normal	[0,1]	0	tcp	http	SF	267	14515
7	normal	[0,1]	0	tcp	smtp	SF	1022	387
8	normal	[0,1]	0	tcp	telnet	SF	129	174
9	normal	[0,1]	0	tcp	http	SF	327	467
10	anomaly	[1,0]	0	tcp	ftp	SF	26	157
11	anomaly	[1,0]	0	tcp	telnet	SF	0	0
12	normal	[0,1]	0	tcp	smtp	SF	616	330
13	anomaly	[1,0]	0	tcp	private	REJ	0	0
14	anomaly	[1,0]	0	tcp	telnet	S0	0	0

CONCLUSION

- ❑ This project successfully demonstrated the creation and deployment of a highly accurate Network Intrusion Detection System.
- ❑ Using IBM Watsonx.ai and its AutoAI capabilities significantly accelerated the development process, automating tasks that would typically require extensive manual coding and expertise.
- ❑ The final deployed model serves as a powerful and scalable solution for enhancing network security through real-time threat detection.

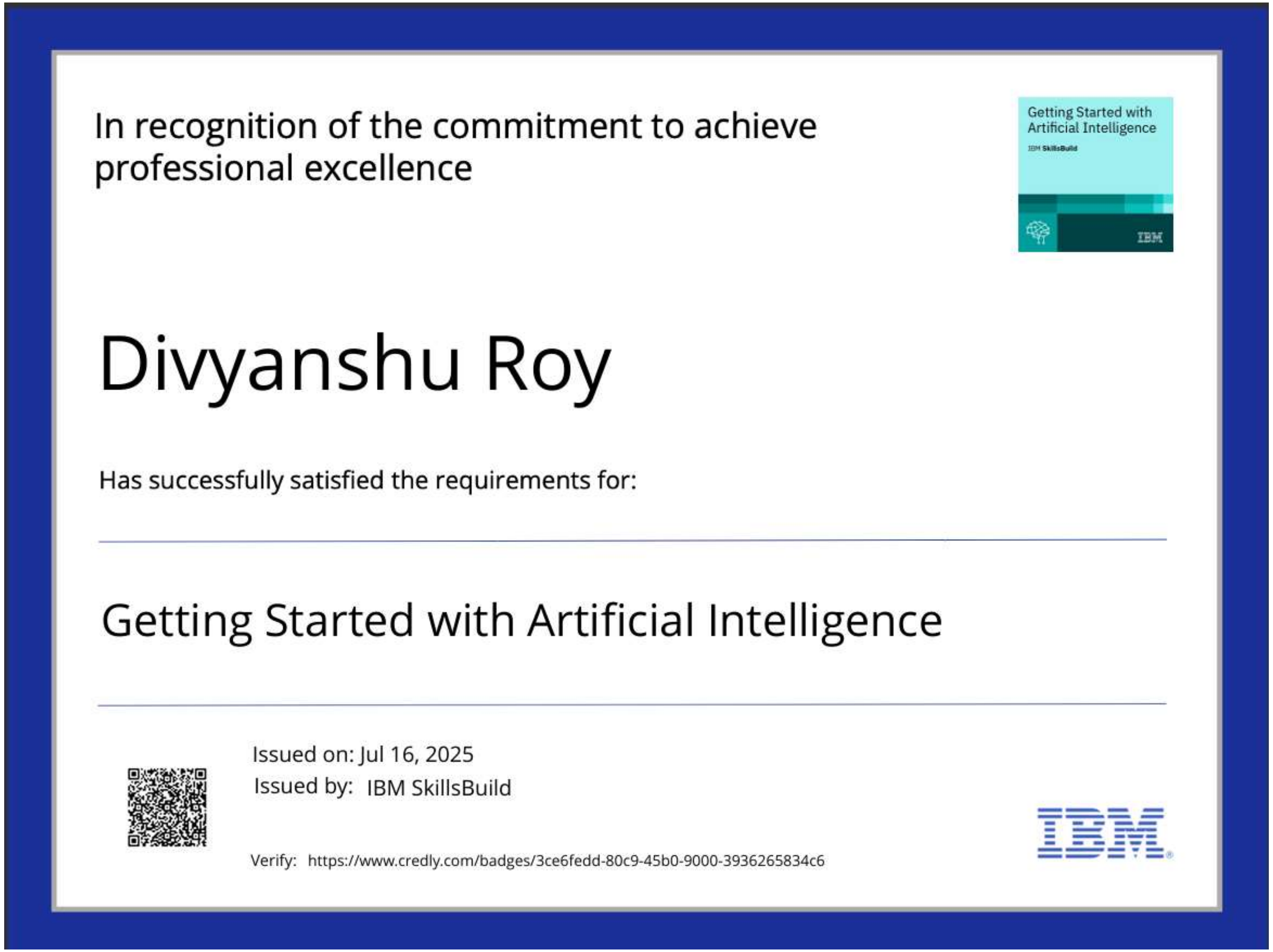
FUTURE SCOPE

- ❑ **Real-time Integration:** Integrate the deployed API with a live network monitoring tool (like Wireshark or a custom dashboard) to analyze traffic in real-time.
- ❑ **Automated Retraining:** Implement a CI/CD pipeline to automatically retrain and redeploy the model as new attack data becomes available.
- ❑ **Advanced Explainability:** Use AI explainability tools to better understand *why* the model flags certain activities as malicious, providing deeper insights for security analysts.

REFERENCES

- ❑ **Dataset:** "NSL-KDD Dataset" from Kaggle.
- ❑ *Link: https://www.kaggle.com/datasets/sampadab17/network-intrusion-detection*
- ❑ **Platform:** IBM Cloud & Watsonx.ai Documentation.

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM SkillsBuild

IBM

Divyanshu Roy

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

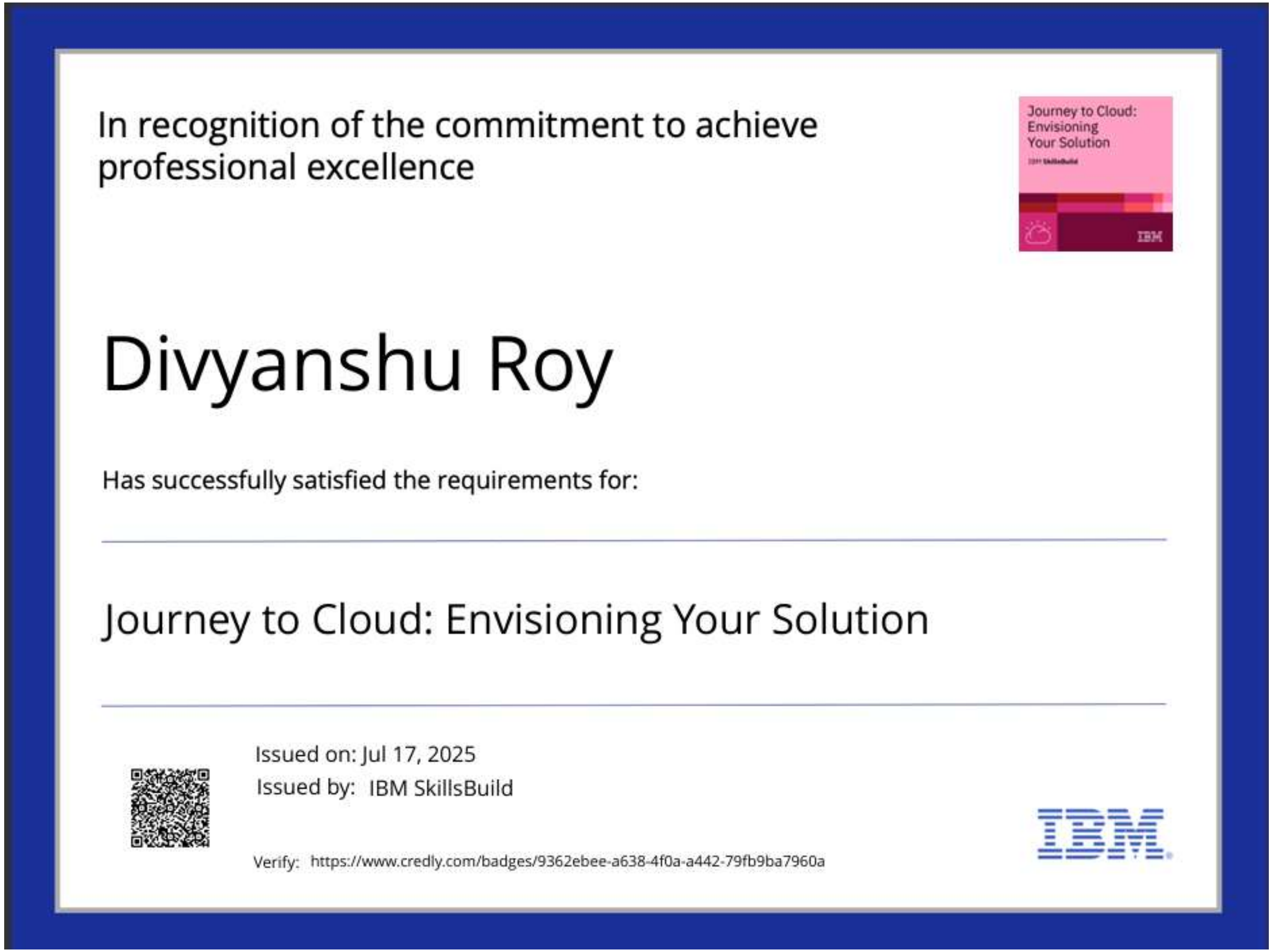
Issued on: Jul 16, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/3ce6fedd-80c9-45b0-9000-3936265834c6

IBM

IBM CERTIFICATIONS


Journey to Cloud:
Envisioning
Your Solution
IBM SkillsBuild
IBM

In recognition of the commitment to achieve professional excellence

Divyanshu Roy

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

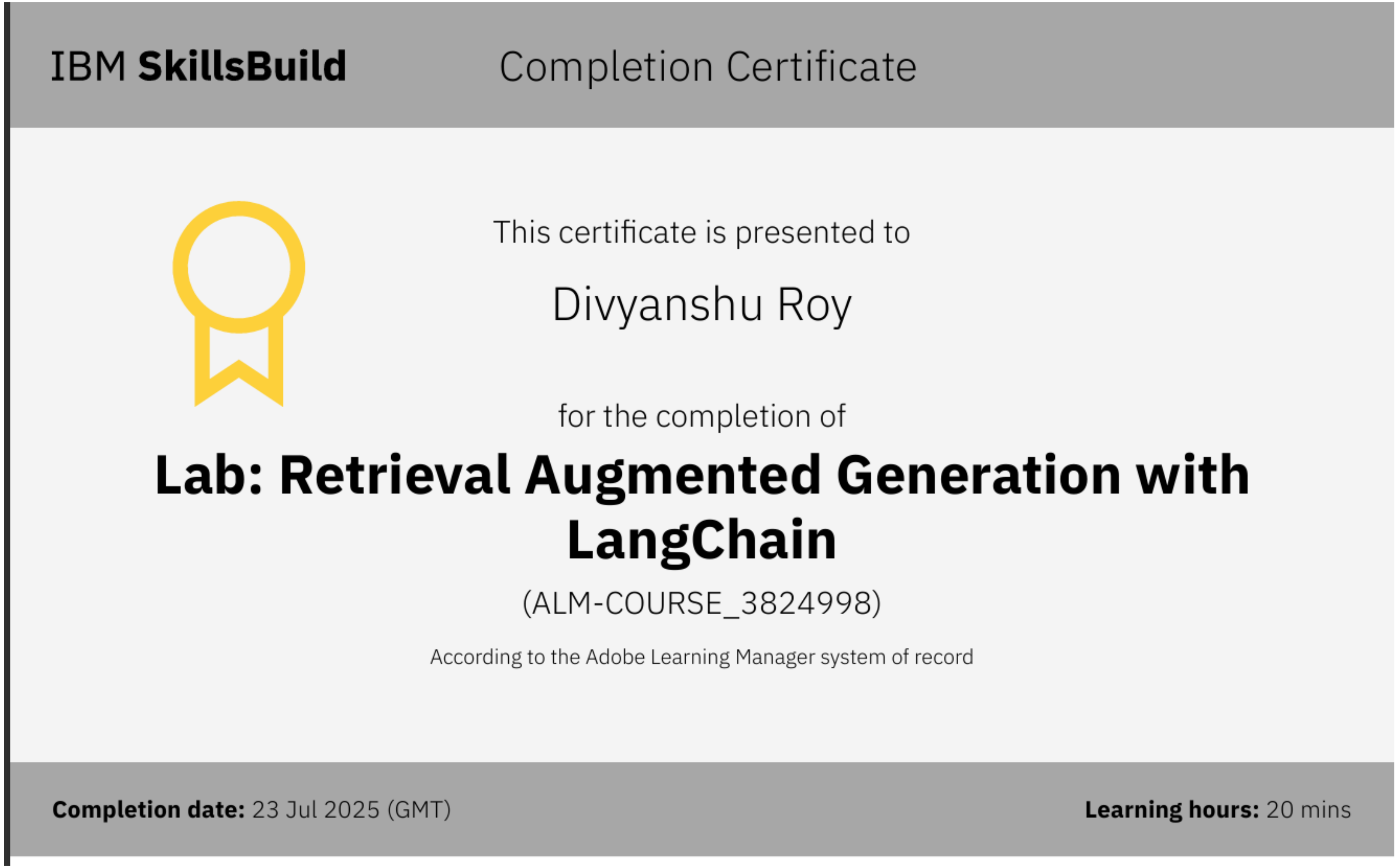
Issued on: Jul 17, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/9362ebee-a638-4f0a-a442-79fb9ba7960a

IBM

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Divyanshu Roy

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU