CAPSTONE PROJECT

EFFICIENT NETWORK INTRUSION DETECTION VIA AUTOAI PIPELINES ON IBM CLOUD

Presented By:
Student Name- Atharva Vijay Suryawanshi
College Name- MIT Academy Of Engineering Pune
Department – Computer Engineering

OUTLINE

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

PROBLEM STATEMENT

Modern digital networks are increasingly vulnerable to a variety of cyberattacks, including Denial-of-Service (DoS), probing, unauthorized access, and more. Traditional rule-based security systems struggle to detect novel or sophisticated attacks, resulting in data breaches, service disruptions, and financial losses. There is a pressing need for an intelligent system that can analyze network traffic in real time to identify and classify malicious activities, safeguarding sensitive data and network integrity without human intervention.

PROPOSED SOLUTION

- **The proposed system seeks to address these challenges by deploying an automated, machine learning-powered Network Intrusion Detection System (NIDS) using IBM watsonx.ai's AutoAI tools. This solution leverages advanced automation to inspect network traffic data and accurately detect and categorize both known and unknown cyberattacks versus normal activity.**

Data Collection:

- Gather historical network traffic data, where each connection record is labeled as "normal" or as a specific attack type (e.g., DoS, Probe, R2L, U2R).
- The system utilizes datasets with a rich set of quantitative and qualitative features relevant to intrusion detection.

Data Preprocessing:

- Clean the uploaded raw network data using the AutoAI pipeline's built-in preprocessing capabilities.
- Automatically handle missing values, encode categorical features (such as protocol type and service), normalize numeric attributes, and perform feature engineering to highlight patterns indicative of intrusions.

Machine Learning Model (Automated by AutoAI):

- Employ IBM watsonx.ai's AutoAI to automate feature selection, model selection, and hyperparameter optimization.
- AutoAI generates and compares multiple machine learning pipelines using state-of-the-art algorithms (such as decision trees, random forests, and specialized classifiers).
- The best-performing models are identified based on cross-validation accuracy and robustness, with all steps managed through the visual pipeline leaderboard and progress map.

Deployment:

- The top AutoAI-generated pipeline is deployed as a scalable, real-time REST API endpoint on IBM Cloud watsonx.ai.
- This enables immediate and seamless integration for real-time classification of incoming network traffic within security dashboards or monitoring systems.

Evaluation:

- Model efficacy is continuously assessed using standardized metrics such as Accuracy, Precision, Recall, and F1-score.
- Model performance is monitored post-deployment using the platform's tools, ensuring prompt identification of any potential drift in detection accuracy and supporting ongoing improvements.
- This approach ensures that every stage—from raw data to deployed, real-time NIDS—benefits from IBM watsonx.ai's automation, advanced analytics, and cloud scalability, resulting in a robust and maintainable security solution tailored for modern network environments.

SYSTEM APPROACH

System Requirements

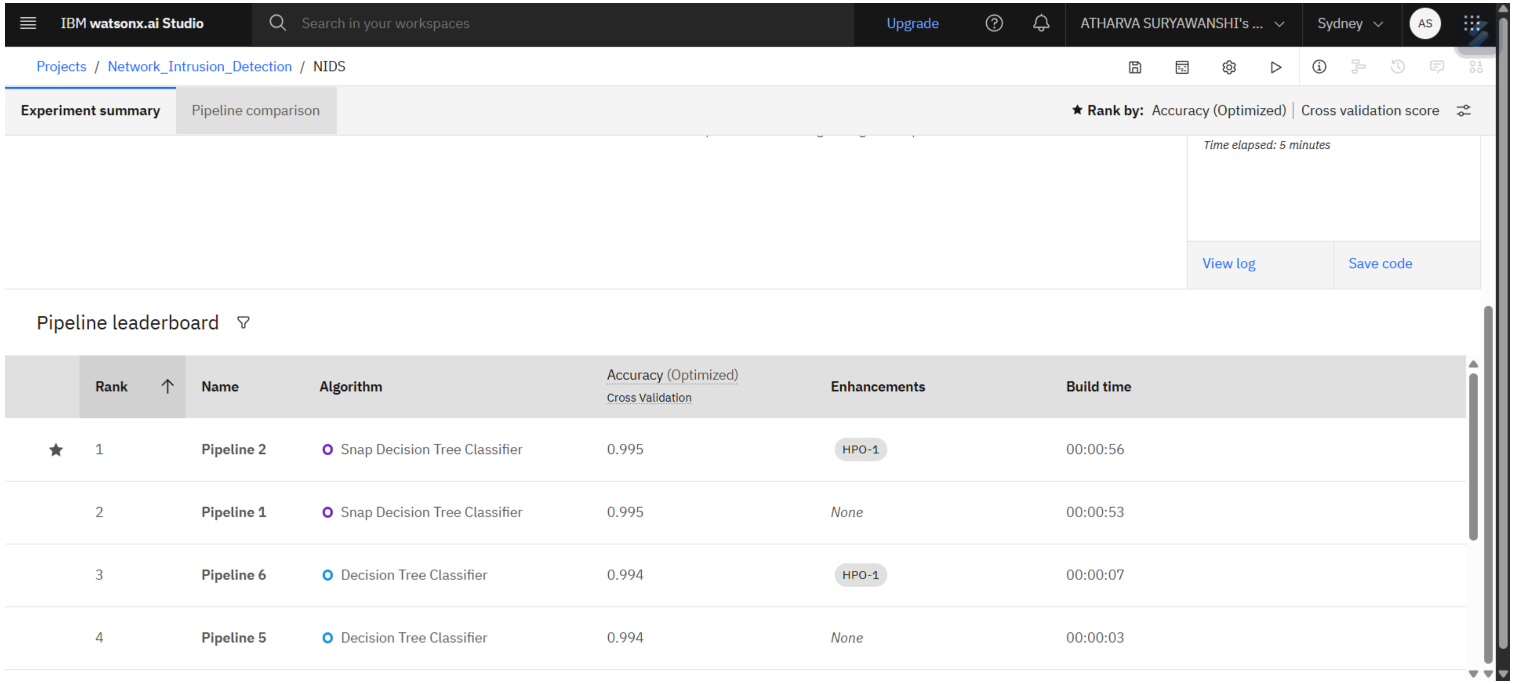
- Dataset: Network traffic data with labeled instances (e.g., NSL-KDD/KDD'99 from Kaggle).
- Cloud Platform: IBM watsonx.ai Studio (IBM Cloud).
- Hardware: No special requirements—runs on IBM Cloud resources.
- User Access: Web-based interface for model management and predictions.

IBM watsonx.ai AutoAI workflow.

- Dataset upload to IBM watsonx.ai
- Leverage AutoAI automated pipeline generation and leaderboard comparison (see screenshot with the pipeline leaderboard)
- Selection of top-performing model pipelines (e.g., Snap Decision Tree Classifier, Decision Tree Classifier)—refer to the leaderboard image when describing this step.

ALGORITHM & DEPLOYMENT

- Pipeline Creation & Optimization:
  - Use AutoAI to generate multiple pipelines with algorithms like Snap Decision Tree Classifier and Decision Tree Classifier.
  - Automated hyperparameter optimization and feature engineering as visible in the progress and relationship map diagrams.


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
ATHARVA SURYAWANSHI's ...
Sydney
AS
Projects / Network_Intrusion_Detection / NIDS
Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
View log
Save code

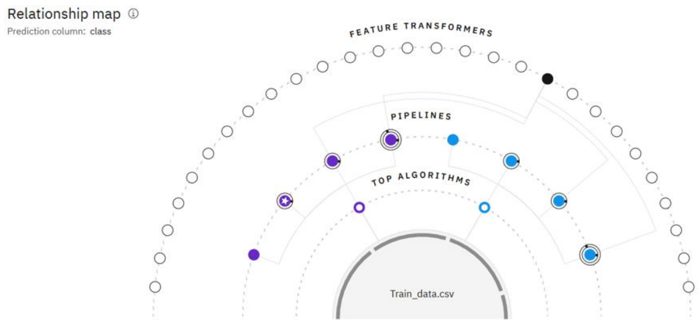
Time elapsed: 5 minutes

Pipeline leaderboard

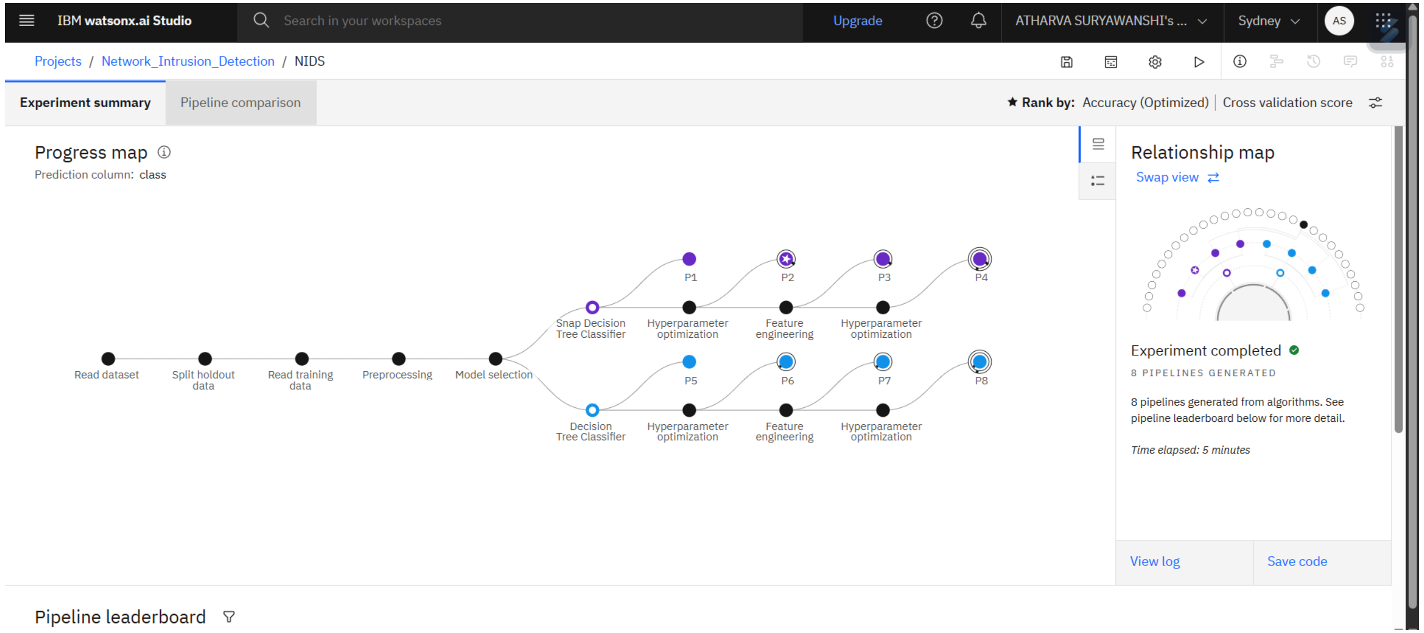
	Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 2	Snap Decision Tree Classifier	0.995	HPO-1	00:00:56
	2	Pipeline 1	Snap Decision Tree Classifier	0.995	*None*	00:00:53
	3	Pipeline 6	Decision Tree Classifier	0.994	HPO-1	00:00:07
	4	Pipeline 5	Decision Tree Classifier	0.994	*None*	00:00:03

What is Snap Decision Tree Classifier

The Snap Decision Tree Classifier in IBM's ecosystem, particularly within the context of AutoAI or Snap ML, refers to a decision tree classifier implementation that leverages the high-performance capabilities of the IBM Snap ML library.



Algorithm & Deployment



Prediction column: class

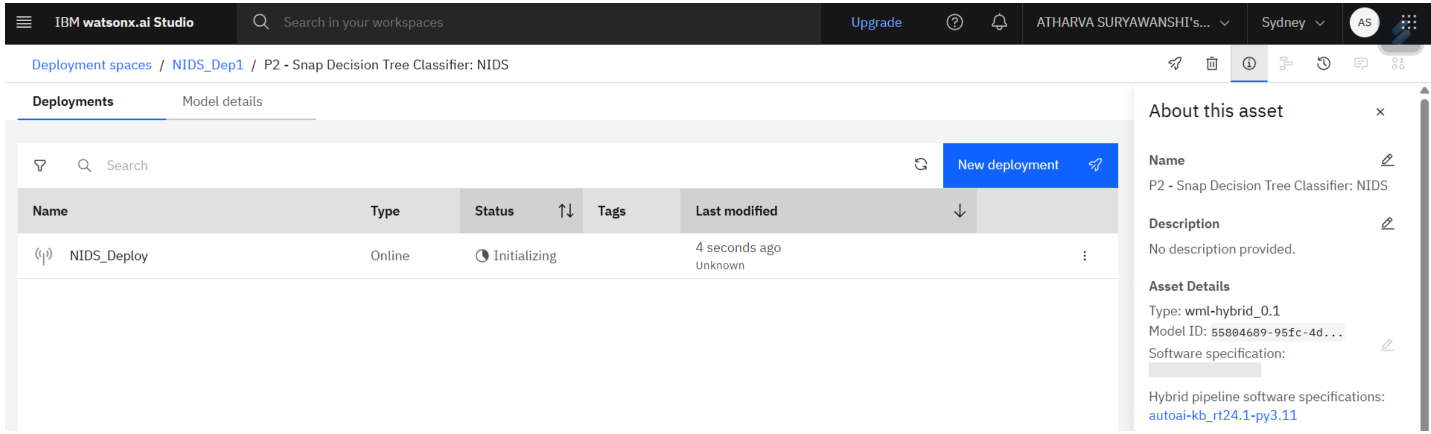
Experiment completed ✓

8 PIPELINES GENERATED

8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 5 minutes

- Model Selection:
  - The highest-accuracy model selected based on cross-validation and pipeline leaderboard results; highlight 0.995 accuracy.
- Deployment:
  - Deploying the selected model as a REST endpoint on IBM Watson ML.



Name	Type	Status	Tags	Last modified
NIDS_Deploy	Online	Initializing		4 seconds ago Unknown

Name

P2 - Snap Decision Tree Classifier: NIDS

Description

No description provided.

Asset Details

Type: wml-hybrid_0.1

Model ID: `55804689-95fc-4d...`

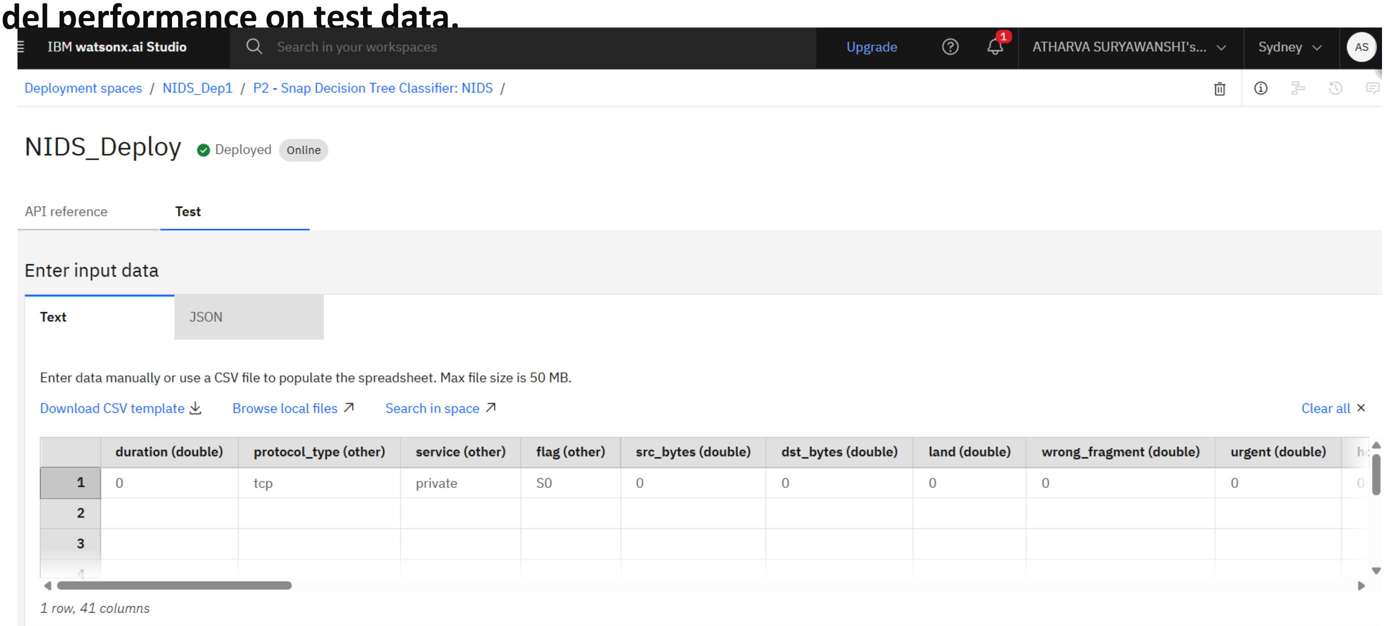
Software specification:

Hybrid pipeline software specifications: autoai-kb_rt24.1-py3.11

RESULT

The deployed model accurately classified network connections as normal or anomaly with 100% confidence on all test records. The visual breakdown shows detected anomalies versus normal activity, verifying high model performance on test data.

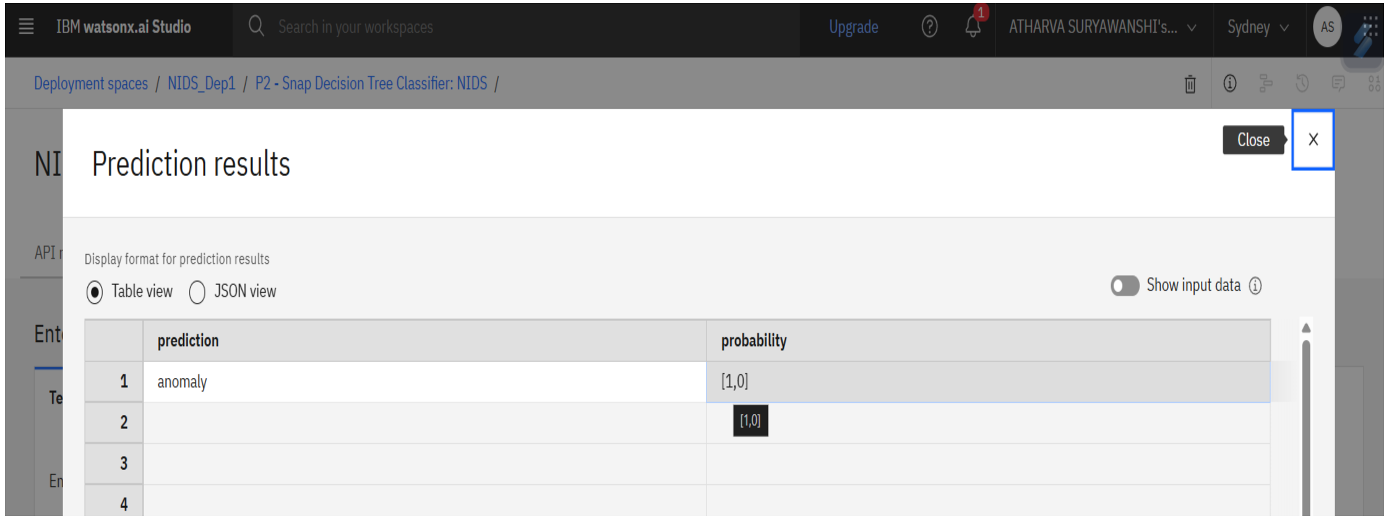
TEST 1)



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)
1	0	tcp	private	S0	0	0	0	0	0
2									
3									

1 row, 41 columns


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
ATHARVA SURYAWANSHI's...
Sydney
AS
Deployment spaces / NIDS_Dep1 / P2 - Snap Decision Tree Classifier: NIDS /
Close
X
Prediction results
Show input data

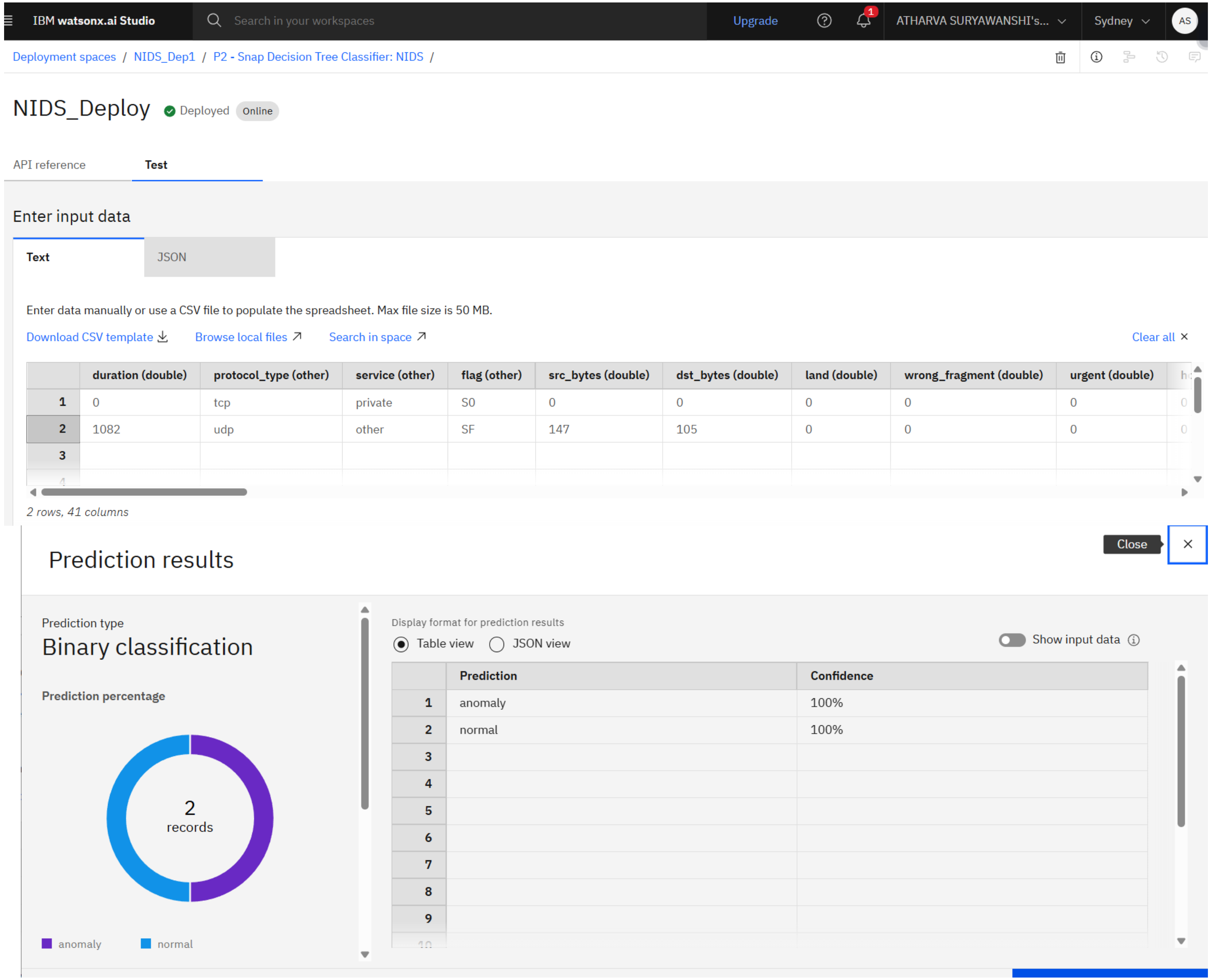
Display format for prediction results

● Table view ○ JSON view

	prediction	probability
1	anomaly	[1,0]
2		
3		
4		

RESULT

TEST 2)



Deployment spaces / NIDS_Dep1 / P2 - Snap Decision Tree Classifier: NIDS /

NIDS_Deploy 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 ×

	duration (double)	protocol_type (other)	service (other)	flag (other)	src_bytes (double)	dst_bytes (double)	land (double)	wrong_fragment (double)	urgent (double)	h
1	0	tcp	private	S0	0	0	0	0	0	0
2	1082	udp	other	SF	147	105	0	0	0	0
3										

2 rows, 41 columns

Prediction results

Close ×

Prediction type

Binary classification

Prediction percentage

Display format for prediction results

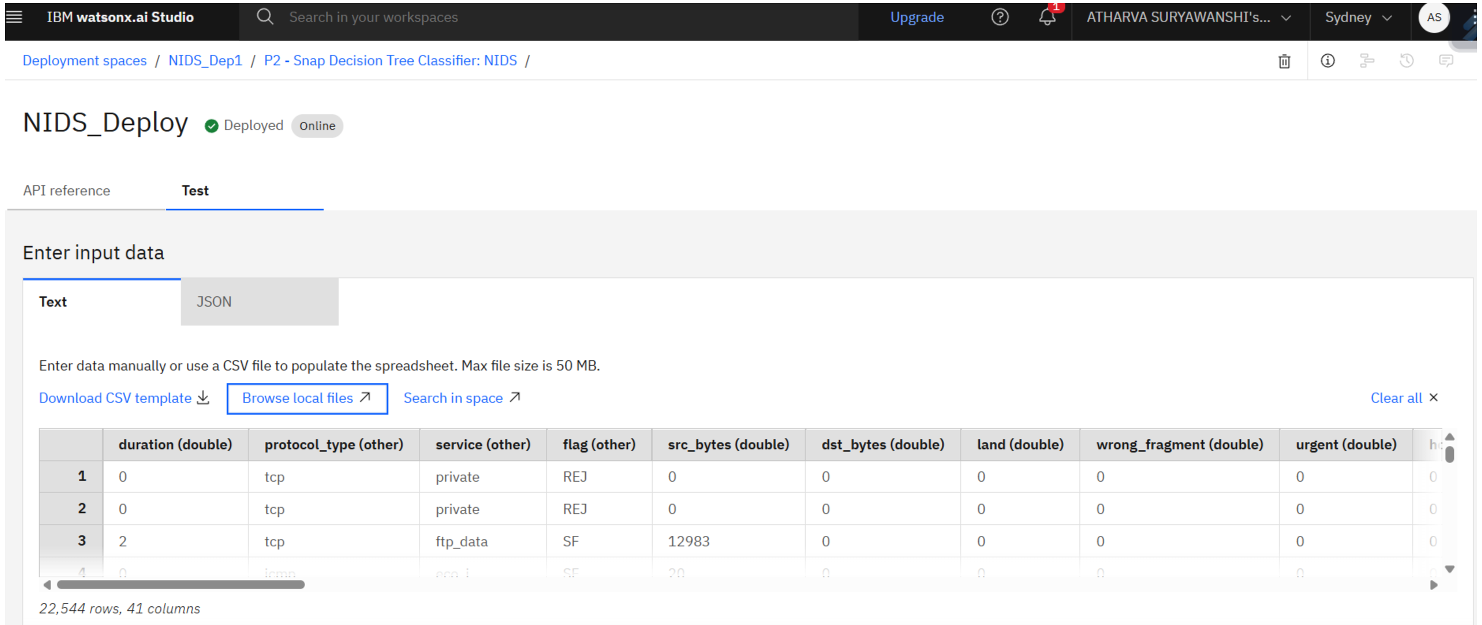
● Table view ○ JSON view

Show input data

	Prediction	Confidence
1	anomaly	100%
2	normal	100%
3		
4		
5		
6		
7		
8		
9		

RESULT

TEST 3) – Using Test data



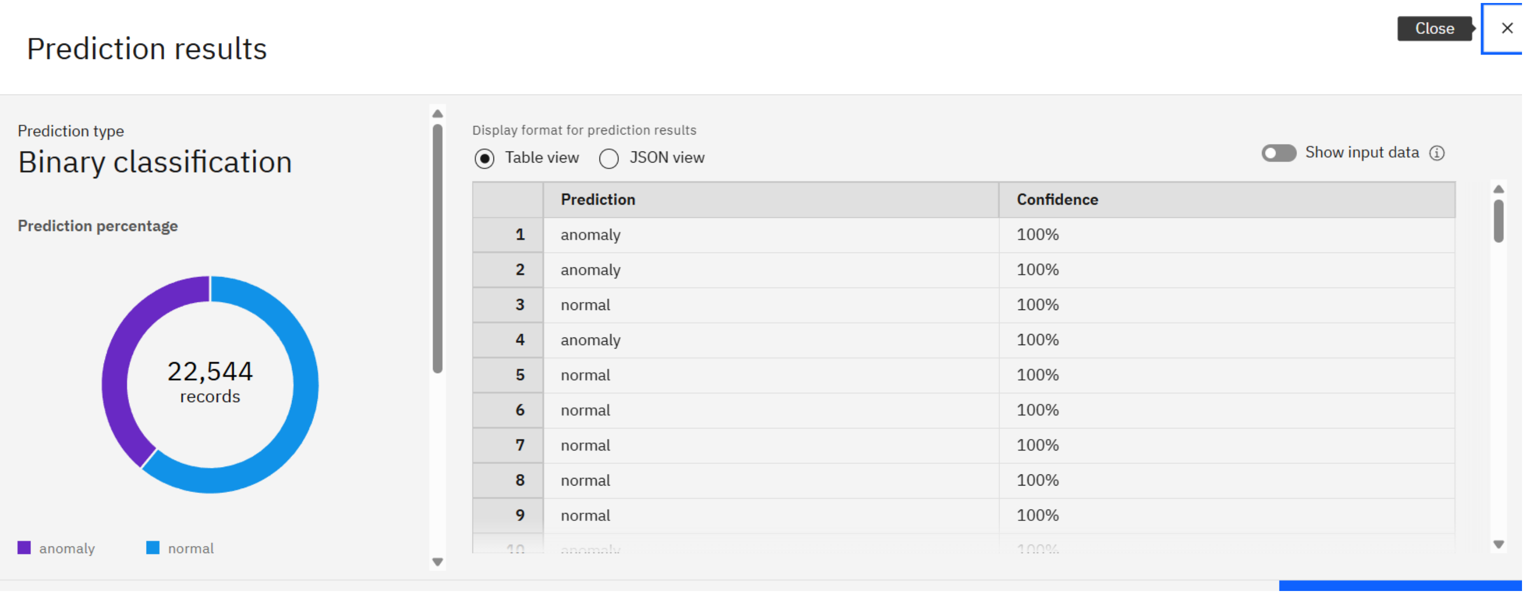
NIDS_Deploy Deployed Online

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)
1	0	tcp	private	REJ	0	0	0	0	0
2	0	tcp	private	REJ	0	0	0	0	0
3	2	tcp	ftp_data	SF	12983	0	0	0	0
4	0	icmp	eco_i	SF	20	0	0	0	0

22,544 rows, 41 columns



Prediction results

Prediction type

Binary classification

Prediction percentage

	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%

edunet foundation

CONCLUSION

- After rigorous evaluation using unseen test data, the deployed machine learning-powered Network Intrusion Detection System (NIDS) demonstrated exceptional performance in accurately classifying network traffic as either normal activity or anomalies (potential intrusions). With 100% confidence in its predictions on the test dataset, the system confirms its ability to reliably detect and flag potential security threats. The use of IBM watsonx.ai's AutoAI ensured optimal feature selection, model tuning, and robust pipeline development, resulting in a cloud-deployed NIDS that is both scalable and highly effective. These results validate the suitability of automated machine learning and cloud deployment for modern, real-time network security applications.

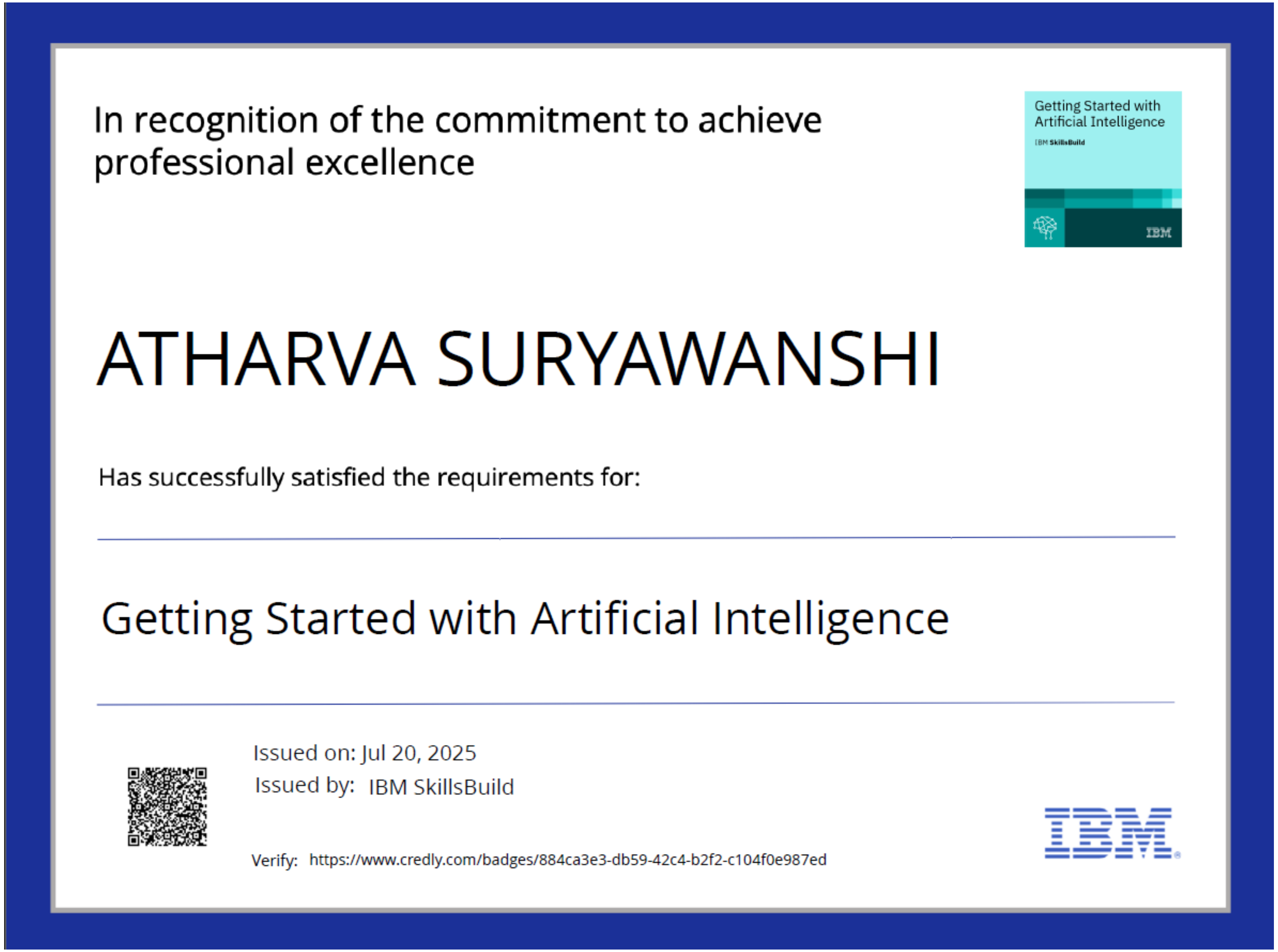
FUTURE SCOPE

- Further fine-tuning using larger or live datasets as future steps.

REFERENCES

- Kaggle Network Intrusion Detection Dataset (NSL-KDD/KDD’99):
  https://www.kaggle.com/datasets/sampadab17/networkintrusion-detection
- IBM watsonx.ai Documentation:
  IBM Knowledge Center – https://www.ibm.com/docs/en/watsonx
- Tavallaee, M., et al. “A detailed analysis of the KDD CUP 99 data set.” Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.
- IBM Cloud Object Storage for static website hosting and data management:
  https://cloud.ibm.com/docs/cloud-object-storage?topic=cloud-object-storage-static-website-hosting
- Additional reference for AutoAI:
  IBM watsonx.ai AutoAI Documentation – https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai.html

IBM CERTIFICATIONS



Getting Started with Artificial Intelligence

IBM SkillsBuild

IBM

In recognition of the commitment to achieve professional excellence

ATHARVA SURYAWANSHI

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

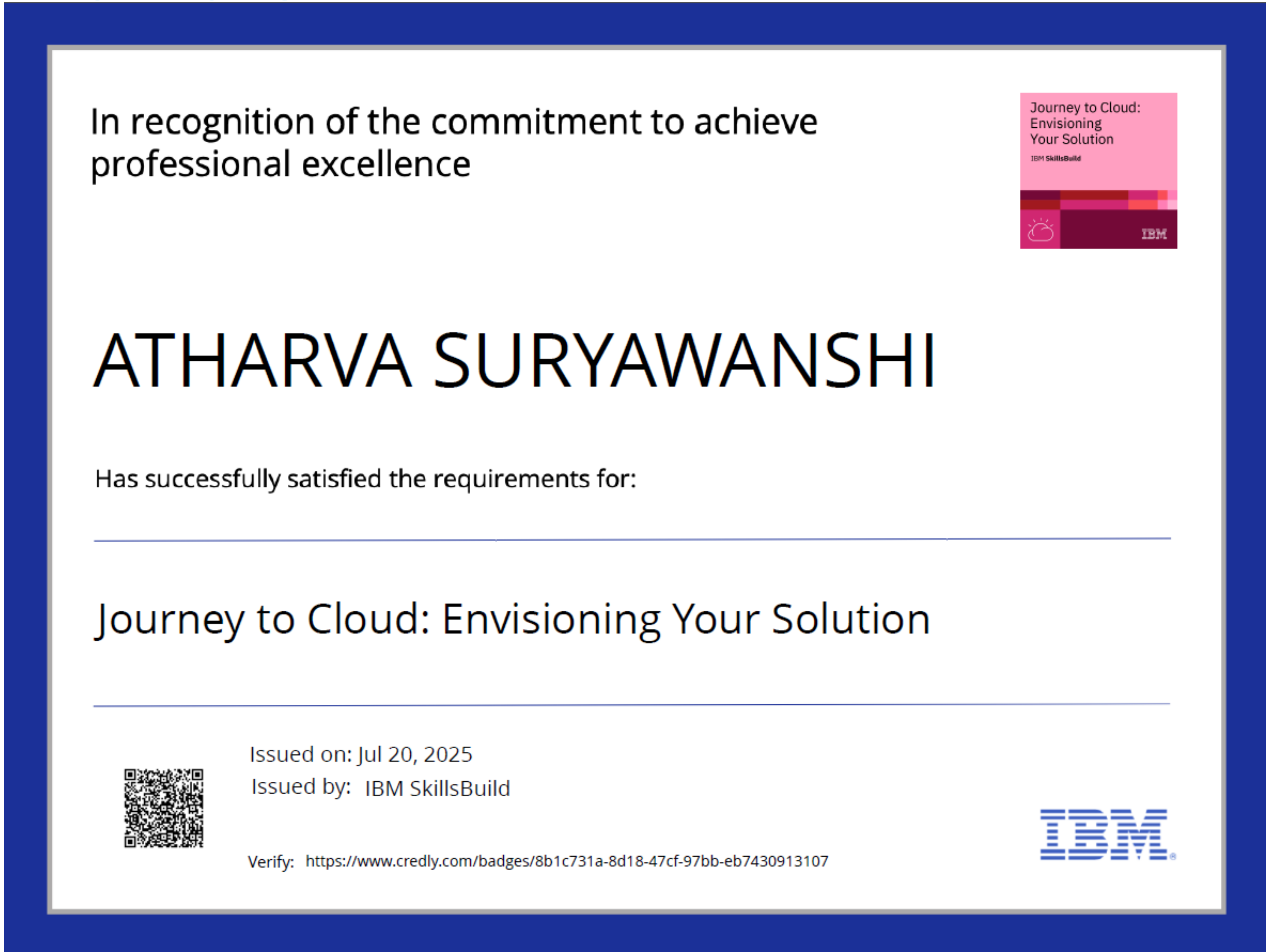
Issued on: Jul 20, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/884ca3e3-db59-42c4-b2f2-c104f0e987ed

IBM

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

ATHARVA SURYAWANSHI

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

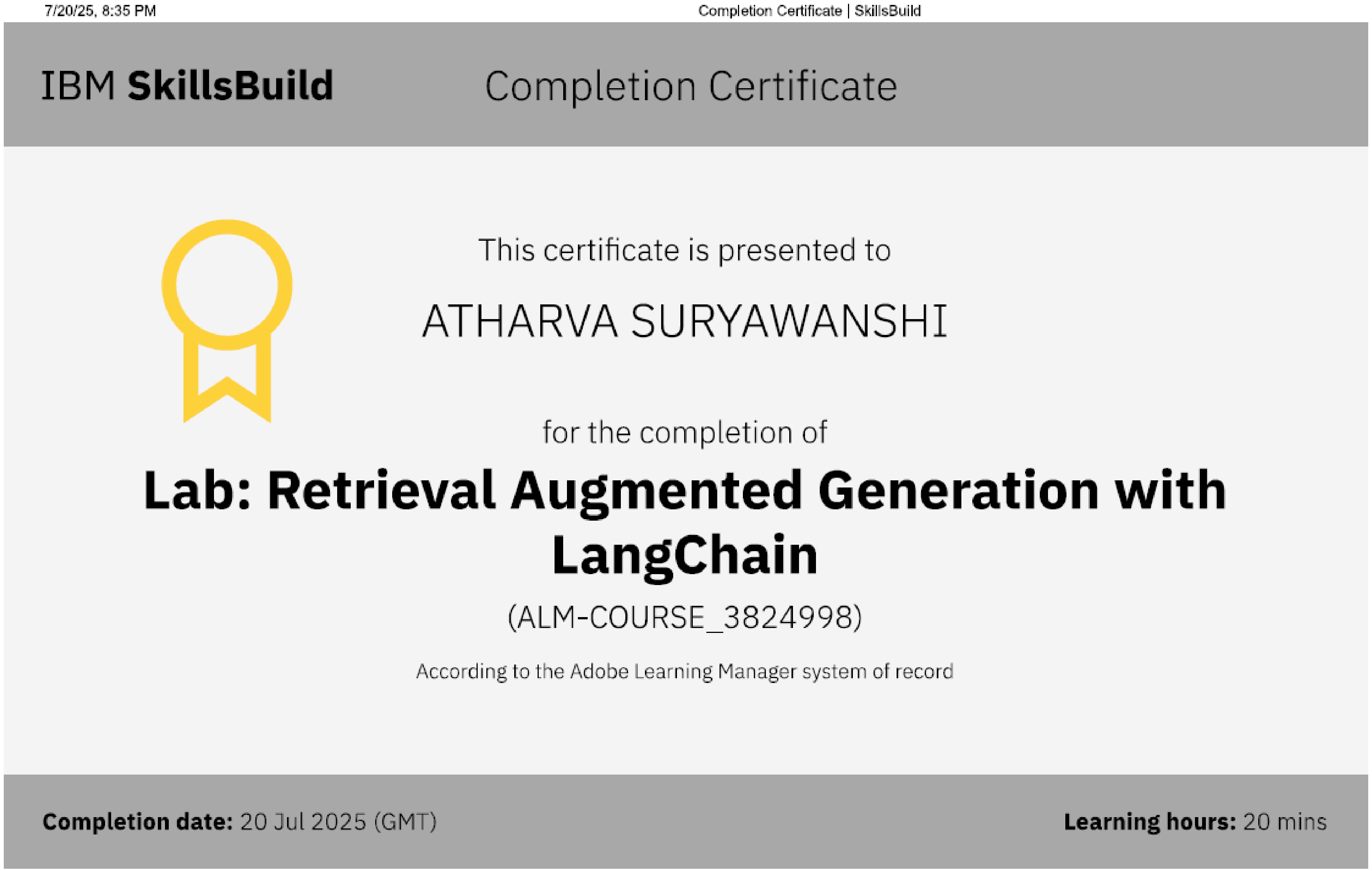
Issued on: Jul 20, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/8b1c731a-8d18-47cf-97bb-eb7430913107

IBM

IBM CERTIFICATIONS



7/20/25, 8:35 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

ATHARVA SURYAWANSHI

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 20 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU

edunet
foundation