CAPSTONE PROJECT

NETWORK INTRUSION DETECTION

Presented By:
Name: Sampad Pal
College: Department of Engineering and Technological Studies ,
University Of Kalyani
Branch: Information Technology

OUTLINE

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

PROBLEM STATEMENT

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.

PROPOSED SOLUTION

To address the network intrusion detection challenge, a machine learning-based model is developed using network traffic data from the Kaggle dataset. The approach includes:

- **Data Collection & Preprocessing**: Cleaning and preparing network traffic features.
- **Feature Engineering**: Transforming and selecting relevant features for higher accuracy.
- **Model Training**: Using IBM Watson AutoAI to explore and optimize classifiers (e.g., Random Forest).
- **Model Evaluation**: Assessing accuracy, precision, recall, and confusion matrix.
- **Deployment**: Deploying the trained model on IBM Cloud Lite using Watson Studio for real-time intrusion detection through an API endpoint.

SYSTEM APPROACH

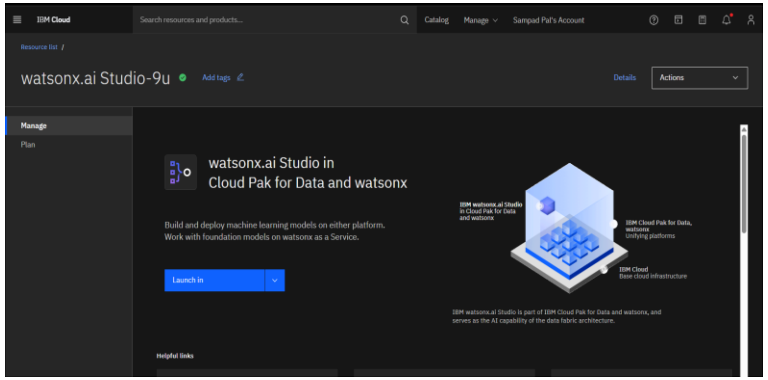
The system development process utilizes IBM Cloud infrastructure to ensure scalability and real-time performance. Key components include:

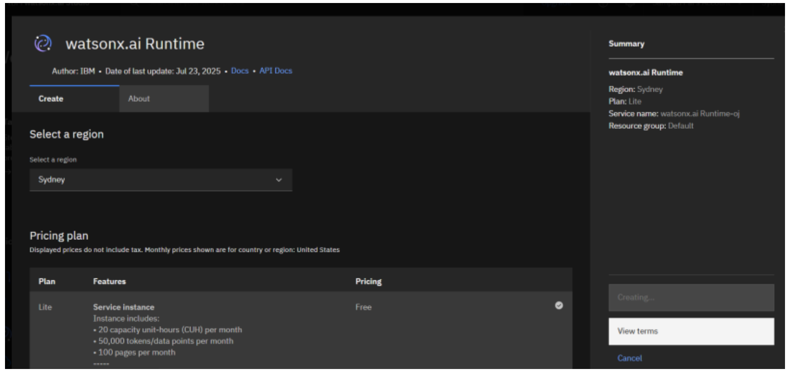
- IBM Watson Studio: For model training and deployment.
- IBM Cloud Object Storage: For securely storing and accessing the dataset.
- IBM Cloud Lite Services: Used to host and deploy the fault detection model via API.

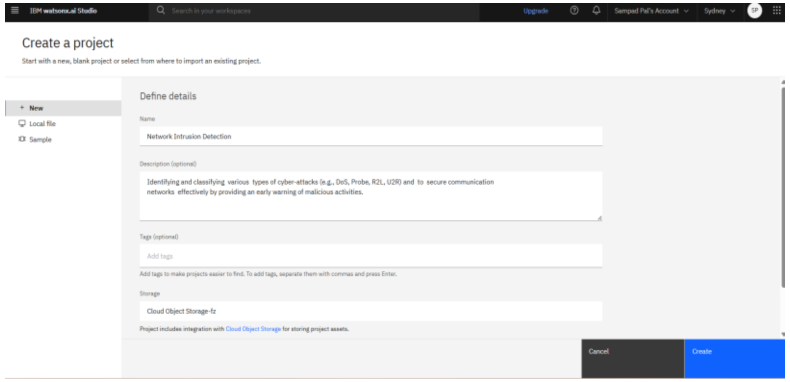
ALGORITHM & DEPLOYMENT

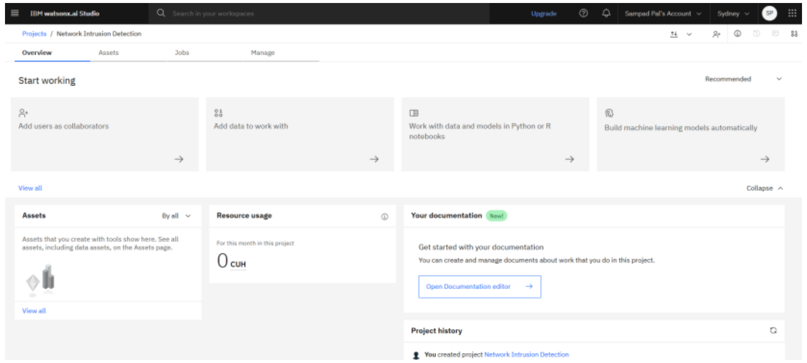
- **Algorithm Used:** Random Forest Classifier (AutoAI optimized using Snap ML)
- **Data Input:** Network traffic features(e.g. protocol_type , src_byte,flag etc)
- **Training Approach:** Supervised Learning using AutoAI in IBM Watson Studio
- **Reason for Selection:** Random Forest is robust, handles non-linear data well, and gives high accuracy for multiclass problems like intrusion detection.
- **Deployment:** The trained model is deployed via **IBM Watson Studio**, providing an API endpoint for real-time predictions and integration into smart grid systems.

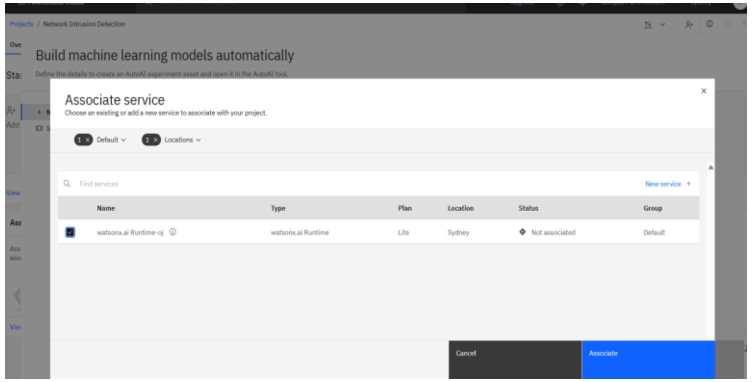
SOME STEPS

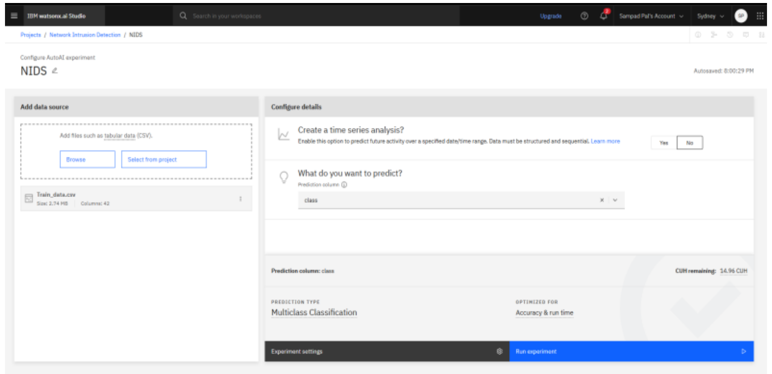






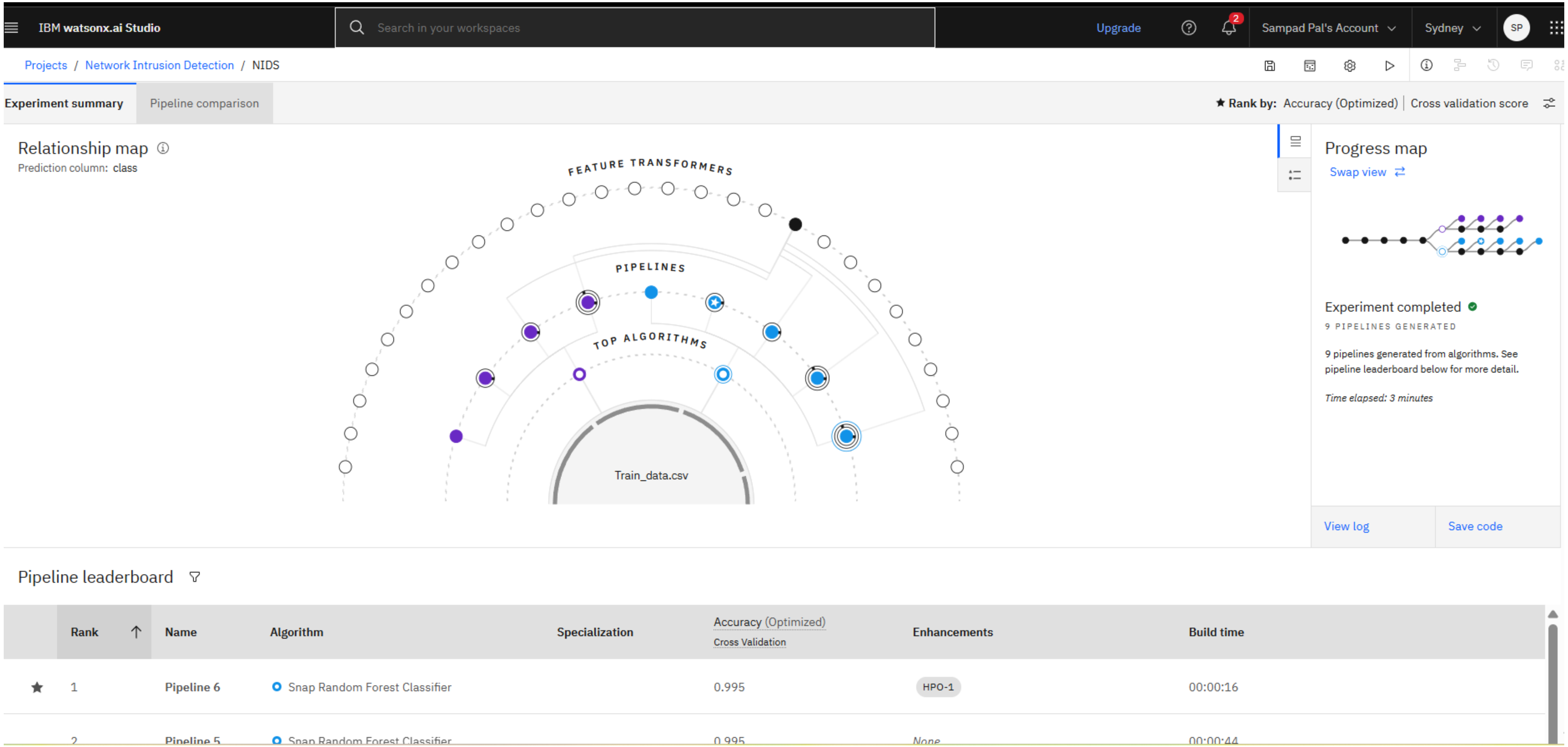




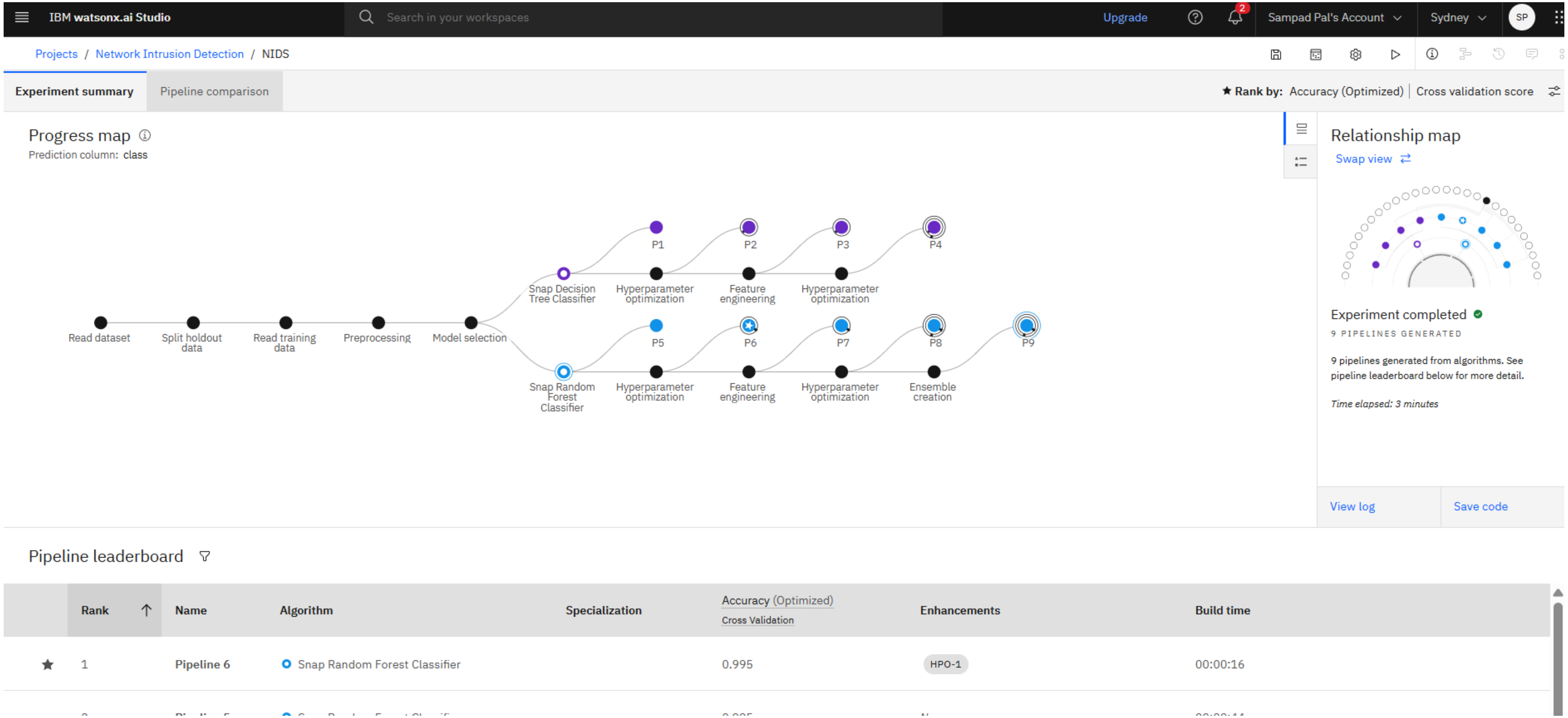


edunet foundation

RESULT

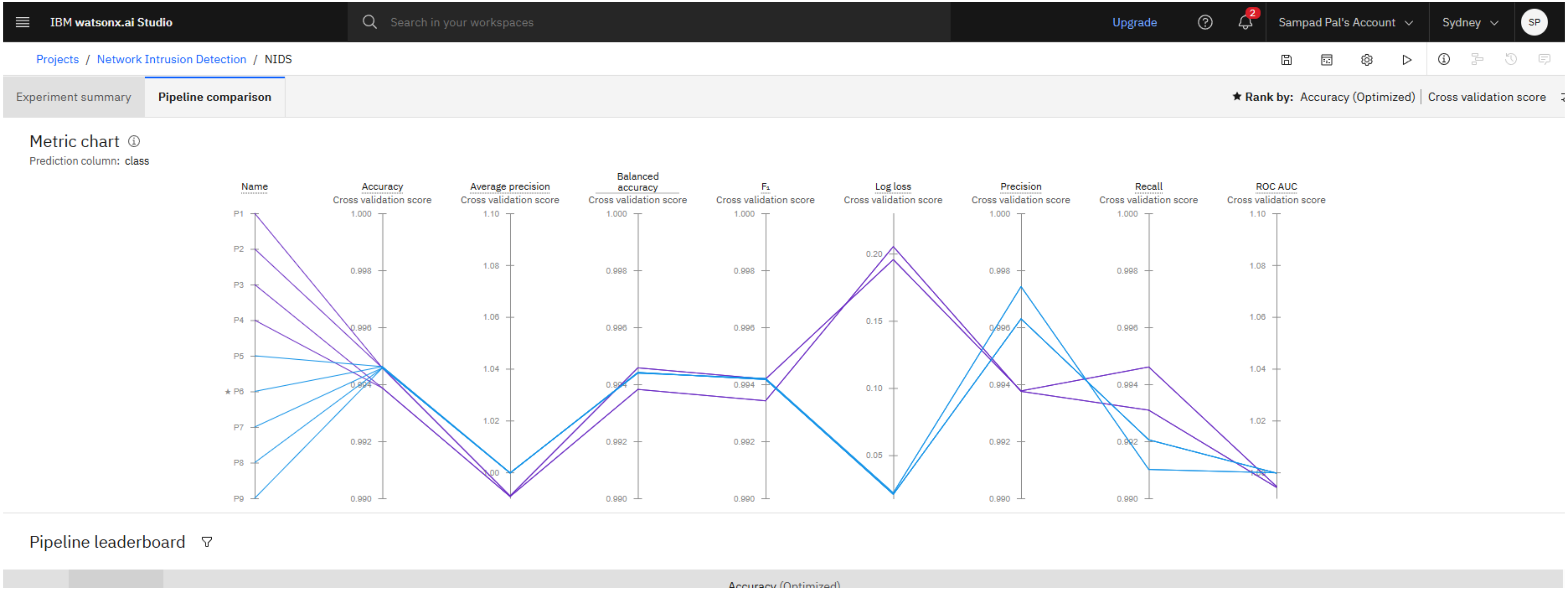

IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Sampad Pal's Account
Sydney
SP
Projects / Network Intrusion Detection / NIDS
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Relationship map
Prediction column: class
FEATURE TRANSFORMERS
PIPELINES
TOP ALGORITHMS
Train_data.csv
Progress map
Swap view
Experiment completed
9 PIPELINES GENERATED
9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 3 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
00:00:16
2
Pipeline 5
Snap Random Forest Classifier
0.995
None
00:00:44

RESULT



IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Sampad Pal's Account
Sydney
SP
Projects / Network Intrusion Detection / NIDS
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Progress map
Prediction column: class
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
Snap Decision Tree Classifier
Snap Random Forest Classifier
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Ensemble creation
P1
P2
P3
P4
P5
P6
P7
P8
P9
Relationship map
Swap view
Experiment completed
9 PIPELINES GENERATED
9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 3 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
00:00:16

RESULT


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Sampad Pal's Account
Sydney
SP
Projects / Network Intrusion Detection / NIDS
Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
Metric chart
Prediction column: class
Name
Accuracy
Cross validation score
Average precision
Cross validation score
Balanced accuracy
Cross validation score
F1
Cross validation score
Log loss
Cross validation score
Precision
Cross validation score
Recall
Cross validation score
ROC AUC
Cross validation score
P1
P2
P3
P4
P5
★ P6
P7
P8
P9
Pipeline leaderboard
Accuracy (Optimized)

RESULT

IBM watsonx.ai Studio | Search in your workspaces | Upgrade | Sampad Pal's Account | Sydney | SP

Deployment spaces / NIDS_Deploy / P6 - Snap Random Forest Classifier: NIDS /

NIDS_Deployment

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)	hot (double)	num_failed_logins (double)	logged
1	0	tcp	smtp	SF	914	329	0	0	0	0	0	1
2	0	tcp	private	S0	0	0	0	0	0	0	0	0
3	0	tcp	smtp	SF	1012	338	0	0	0	0	0	1
4	0	tcp	http	SF	243	667	0	0	0	0	0	1
5	0	tcp	http	SF	227	286	0	0	0	0	0	1
6	0	tcp	private	SH	0	0	0	0	0	0	0	0
7	781	tcp	http	RSTR	79864	0	0	0	0	0	0	1
8	0	tcp	http	SF	259	6391	0	0	0	0	0	1
9	0	tcp	http	SF	233	330	0	0	0	0	0	1

10 rows, 41 columns

Predict

edunet foundation

RESULT

Prediction results

×

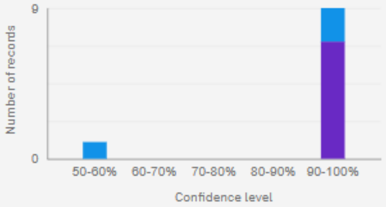
Prediction type

Binary classification

Prediction percentage

normal anomaly

Confidence level distribution



normal anomaly

Display format for prediction results

Table view JSON view

Show input data

	Prediction	Confidence	duration	protocol_type	service	flag	src_bytes	dst_bytes	land
1	normal	99%	0	tcp	smtp	SF	914	329	0
2	anomaly	100%	0	tcp	private	S0	0	0	0
3	normal	100%	0	tcp	smtp	SF	1012	338	0
4	normal	100%	0	tcp	http	SF	243	667	0
5	normal	100%	0	tcp	http	SF	227	286	0
6	anomaly	100%	0	tcp	private	SH	0	0	0
7	anomaly	51%	781	tcp	http	RSTR	79864	0	0
8	normal	100%	0	tcp	http	SF	259	6391	0
9	normal	100%	0	tcp	http	SF	233	330	0
10	normal	100%	0	tcp	http	SF	235	1075	0
11									
12									
13									
14									
15									
16									

CONCLUSION

- The proposed NIDS effectively classifies network traffic into normal and malicious categories using supervised machine learning. Leveraging IBM Watson AutoAI, the system automates feature selection, model tuning, and deployment. The Random Forest model achieved high classification performance, demonstrating its capability in early intrusion detection.
- The use of cloud deployment ensures scalability, easy access, and real-time integration with existing network systems. Challenges faced included dealing with class imbalance and high-dimensional feature space, which were mitigated by AutoAI's automated preprocessing pipeline.

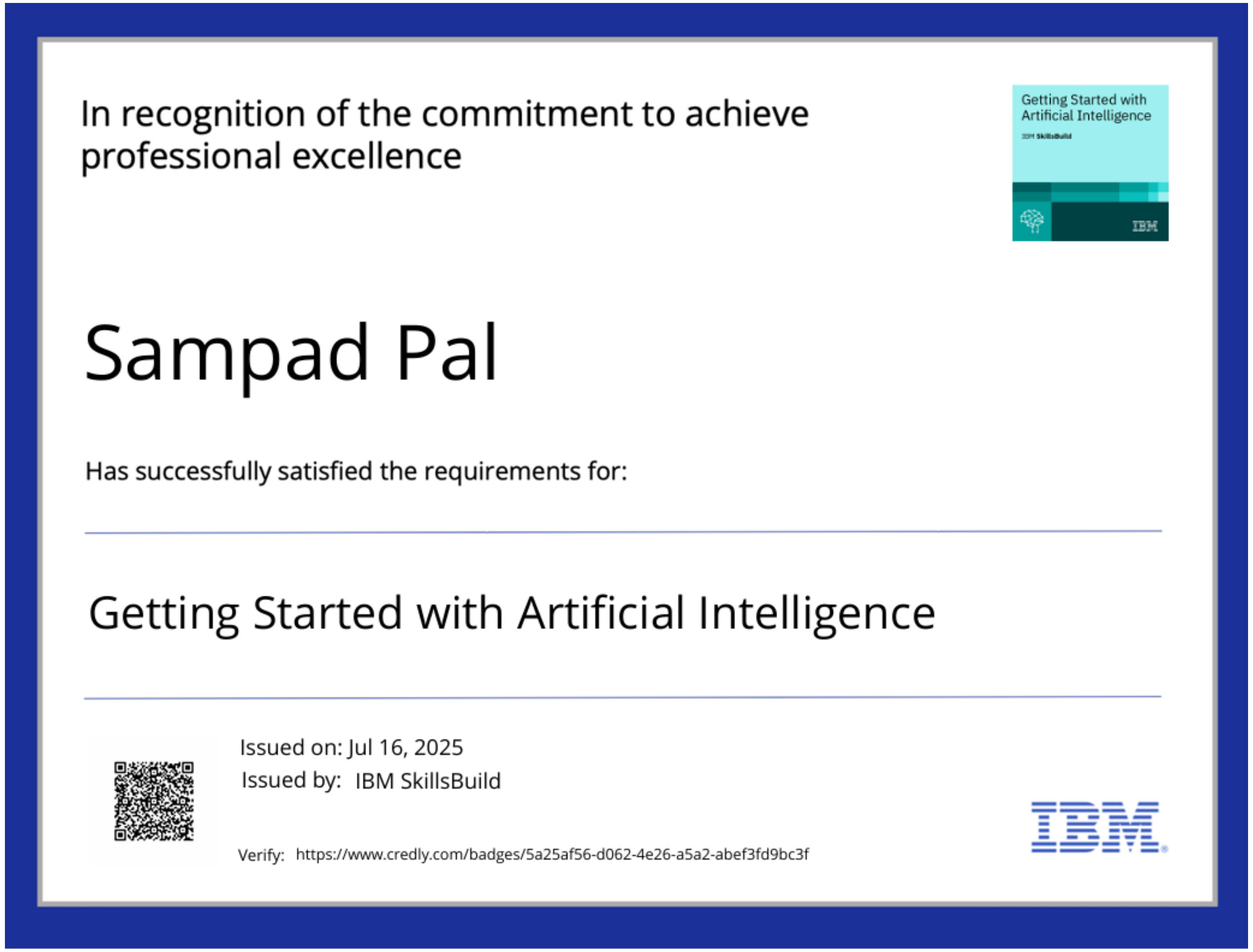
FUTURE SCOPE

- **Class-Specific Detection:** Extend model to separately classify attack types (DoS, Probe, R2L, U2R).
- **Deep Learning Integration:** Use advanced models like LSTM or autoencoders for anomaly detection in sequential network data.
- **Real-Time Monitoring:** Integrate the deployed model with network sensors for live threat detection dashboards.
- **Edge Computing:** Deploy lightweight versions of the model on IoT/edge devices for low-latency detection.
- **Adaptive Learning:** Implement online learning to update the model with new attack types as they evolve.

REFERENCES

- IBM Documentation – https://www.ibm.com/products/watson-studio
- Kaggle Dataset – https://www.kaggle.com/datasets/sampadab17/network-intrusion-detection

IBM CERTIFICATIONS



Getting Started with Artificial Intelligence

In recognition of the commitment to achieve professional excellence

Sampad Pal

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/5a25af56-d062-4e26-a5a2-abef3fd9bc3f

IBM

IBM CERTIFICATIONS

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Sampad Pal

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

GitHub [dapmaS-dev/Network-Intrusion-Model]

THANK YOU

edunet foundation