CAPSTONE PROJECT

NIDS DETECTION ML

Presented By:
Ambarapu Humaun Basha
- **Sathyabama Institute of Science & Technology**
- **Computer Science and Engineering**

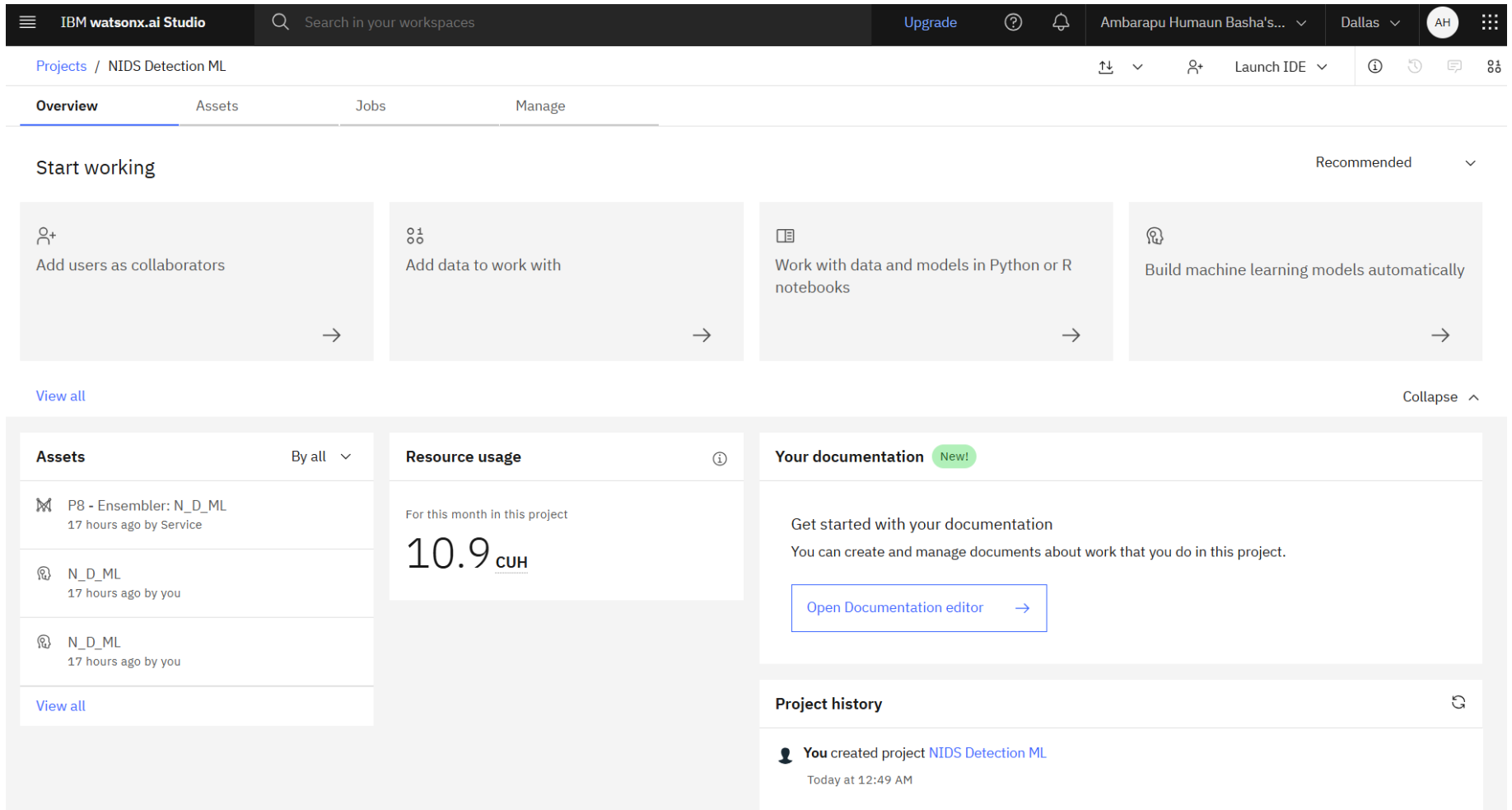
edunet foundation

OUTLINE

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

PROBLEM STATEMENT

Communication networks are increasingly vulnerable to cyber-attacks such as Denial of Service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R). These threats can disrupt services, steal data, and cause significant damage. The challenge is to analyze network traffic data to accurately identify and classify these attacks while distinguishing them from normal activity. Manual detection is inefficient, making an automated system essential for early warnings and network security.



PROPOSED SOLUTION

- The proposed system aims to address the challenge of network intrusion detection by developing a machine learning-based NIDS. This involves analyzing network traffic data to classify attacks and normal activity accurately. The solution will consist of the following components:
- Data Collection:
  - Gather network traffic data from the Kaggle dataset, including features like duration, src_bytes, dst_bytes, and attack types.
  - Utilize additional factors such as traffic patterns and exogenous features for enhanced accuracy.
- Data Preprocessing:
  - Clean and preprocess data to handle missing values, outliers, and inconsistencies.
  - Feature engineering to extract relevant attributes impacting intrusion detection.
- Machine Learning Algorithm:
  - Implement an ensemble model (automatically selected by IBM AutoAI) to classify intrusions based on traffic patterns.
  - Incorporate factors like duration, bytes transferred, and login attempts for improved accuracy.
- Deployment:
  - Develop a scalable interface for real-time intrusion predictions.
  - Deploy on IBM Cloud Lite for reliability and accessibility.
- Evaluation:
  - Assess performance using metrics like SMAPE (Symmetric Mean Absolute Percentage Error).
  - Fine-tune based on pipeline evaluations and monitoring.

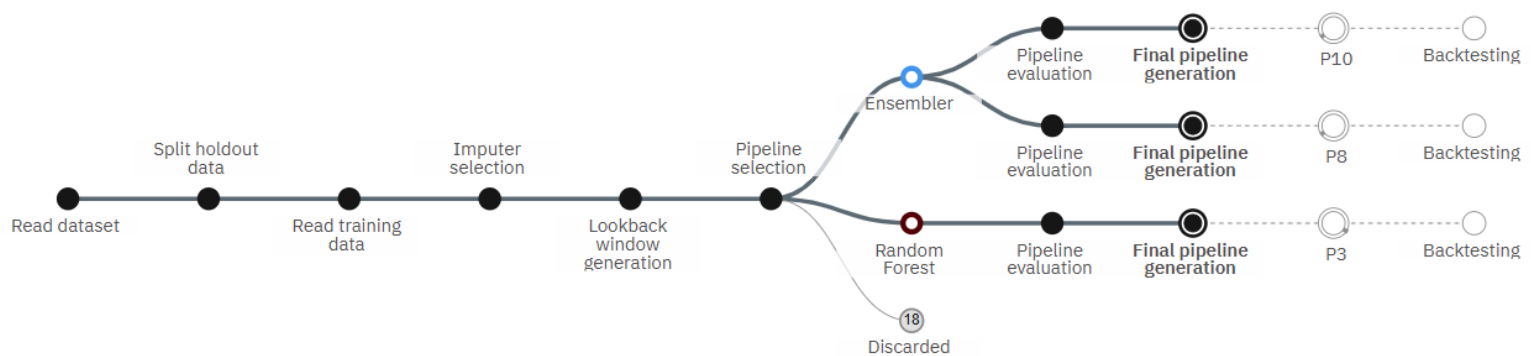
Projects / NIDS Detection ML / N_D_ML

Experiment summary | Pipeline comparison

★ Rank by: Symmetric mean absolute percent... | Backtest score

Progress map

Prediction column: src_bytes



Relationship map

Swap view

Running

Composing pipeline P10

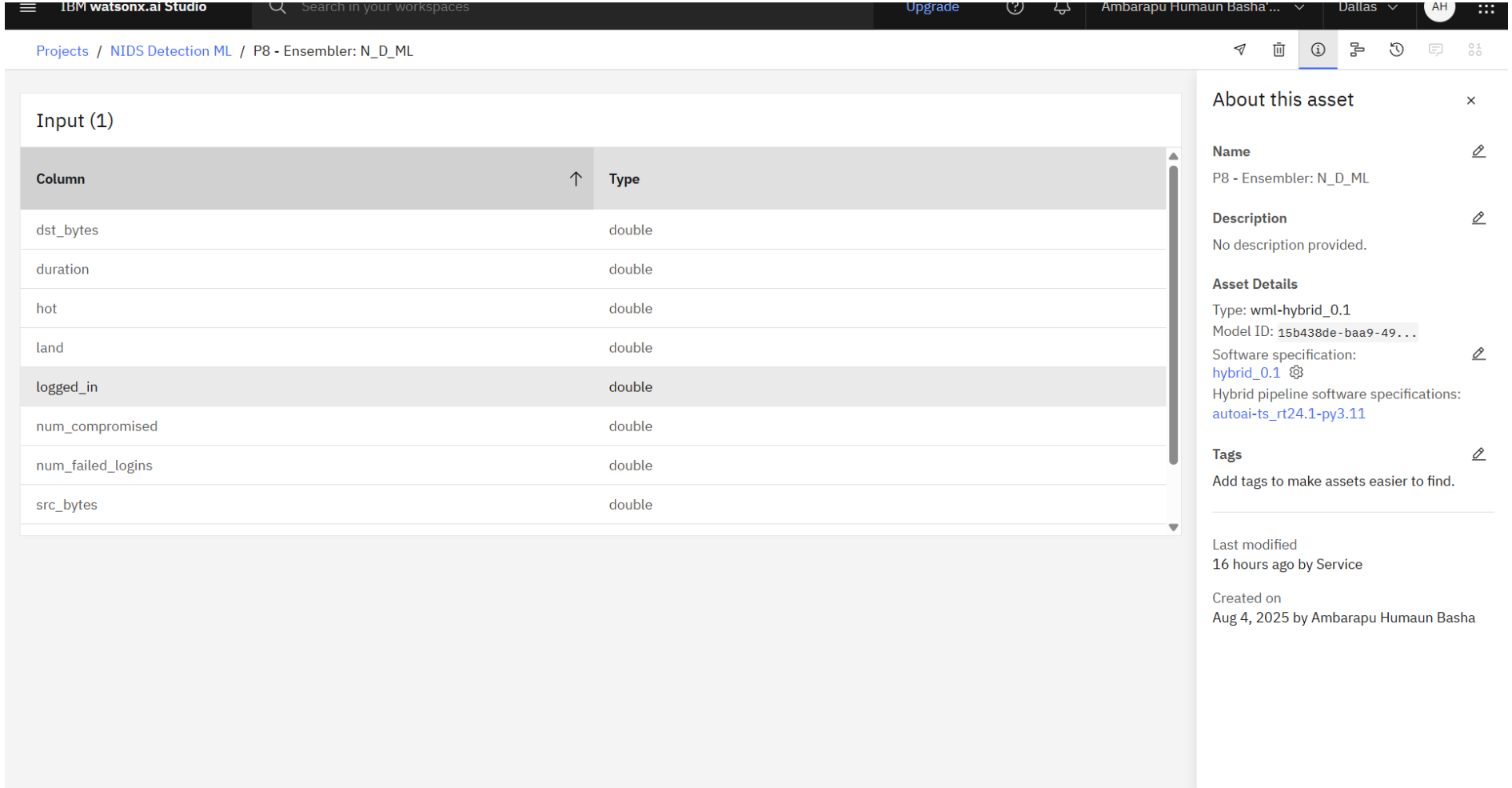
Time elapsed: 20 minutes

View log | Save code

Pipeline leaderboard

SYSTEM APPROACH

- The "System Development Approach" outlines the overall strategy and methodology for developing and implementing the NIDS.
- **System Requirements:**
- IBM Cloud Lite services for model building and deployment.
- Access to datasets and computational resources.
- **Libraries Required to Build the Model:**
- Pandas and NumPy for data handling.
- Scikit-learn for preprocessing and evaluation.
- IBM AutoAI libraries for automated pipeline generation.
- Technology Used: IBM Watsonx.ai Studio for automated ML, ensuring mandatory use of cloud services.



ALGORITHM & DEPLOYMENT

- Algorithm Selection:
  - IBM AutoAI automatically selected the Ensembler algorithm (P8-Ensembler) for its effectiveness in classifying network intrusions, based on data characteristics.
- Data Input:
  - Input features: src_bytes, dst_bytes, duration, land, wrong_fragment, urgent, hot, num_failed_logins, logged_in, etc.
  - Target: Classification of attacks (e.g., DoS, Probe) vs. normal activity.
- Training Process:
  - Trained using historical data with automated backtesting, hyperparameter optimization (HPO), feature engineering (FE), and supplementary techniques (SUP).
- Prediction Process:
  - The model predicts intrusions in real-time, using exogenous features for forecasts (e.g., 1-step ahead).
- Deployment: Deployed as "NIDS Deployment" in IBM Watsonx.ai space, with API endpoints for scoring.

Pipeline leaderboard

	Rank	Name	Algorithm	SMAPE (Optimized) Validation	SMAPE (Optimized) Holdout	SMAPE (Optimized) Backtest	Enhancements	Build time
★	1	**Pipeline 8**	Ensembler	103.405	102.876	102.616	HPO FE SUP	00:00:04
	2	**Pipeline 3**	Random Forest	112.715	104.651	110.670	HPO FE SUP	00:10:19
	3	**Pipeline 10**	Ensembler	126.499	139.460	131.059	HPO FE SUP	00:01:40

Deployment spaces / NDIS detection Deployment space / P8 - Ensembler: N_D_ML

Deployments | Model details

Search | New deployment

Name	Type	Status	Tags	Last modified
NIDS Deployment	Online	Deployed		1 minute ago Ambarapu Humaun Basha (You)

Items per page: 20 | 1–1 of 1 items | 1 of 1 pages

About this asset

Name
P8 - Ensembler: N_D_ML

Description
No description provided.

Asset Details
Type: wml-hybrid_0.1
Model ID: `3c47b514-bfac-43...`
Software specification:
hybrid_0.1
Hybrid pipeline software specifications:
autoai-ts_rt24.1-py3.11

Tags
Add tags to make assets easier to find.

Source asset details

Last modified
2 minutes ago by Service

Created on
Aug 4, 2025 by Ambarapu Humaun Basha

Enter input data

Text | JSON

Request a prediction without new observations, or enter new observations manually or use CSV file to populate the spreadsheet. Max file size is 50 MB.

Forecast window ⓘ

1 – + step ahead

Request a prediction on new observation data. Enter or upload values for the prediction columns and the supporting columns, then click Predict to see the new predictions.

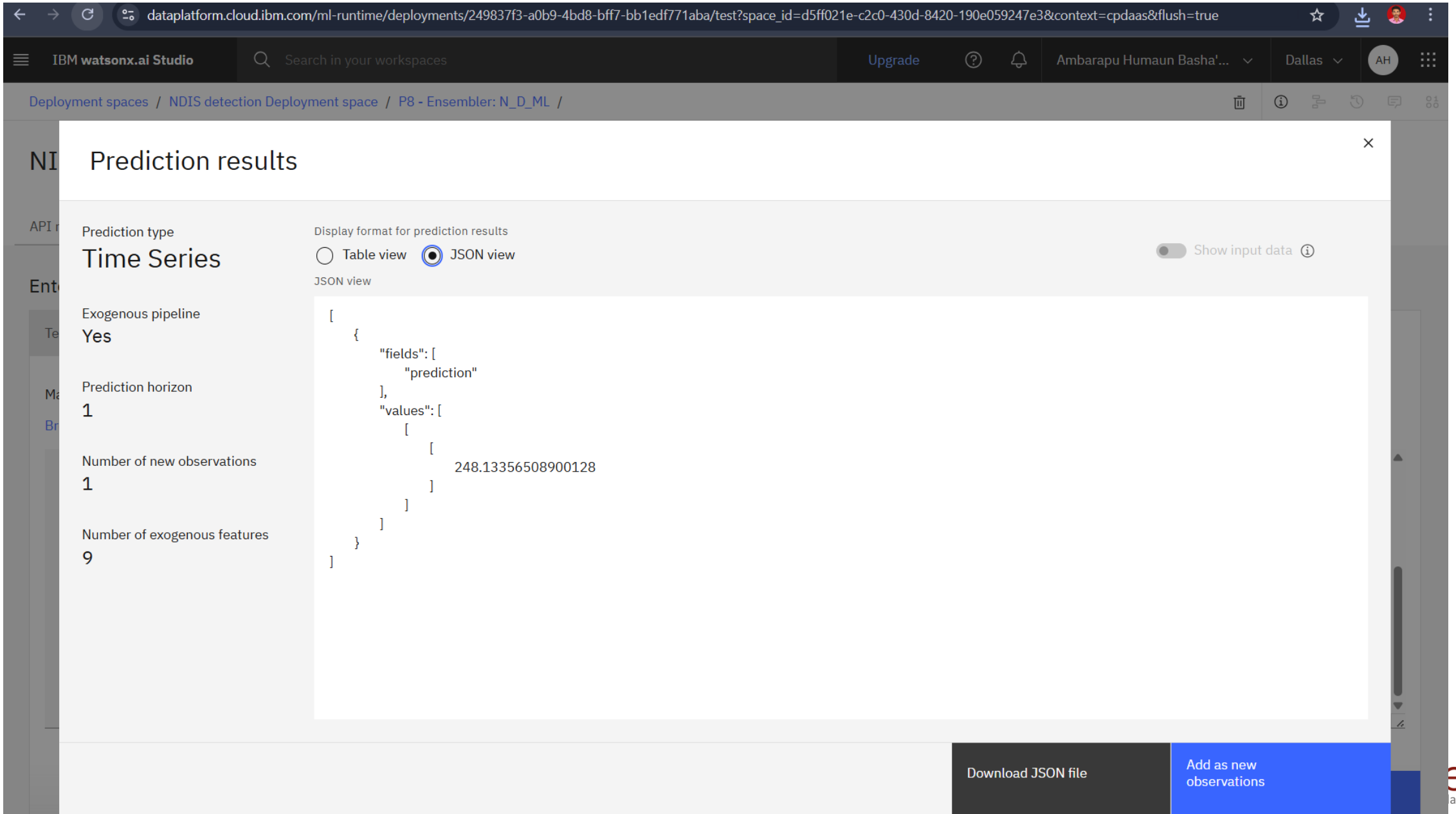
Download CSV template ↓ Browse local files ↗ Search in space ↗ Clear all ×

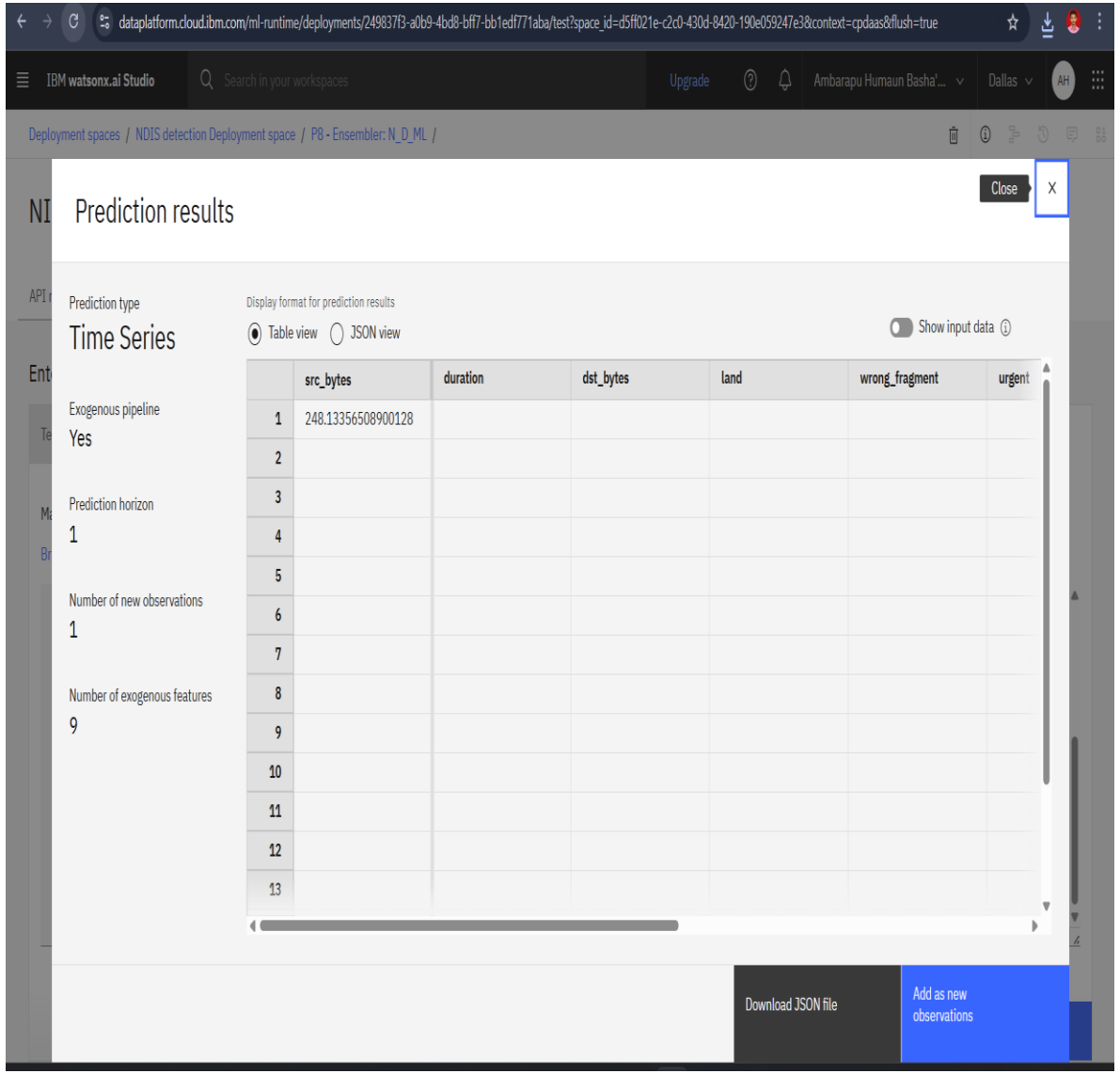
	src_bytes (double)	duration (double)	dst_bytes (double)	land (double)	wrong_fragment (double)	urgent (double)	hot (double)	num_failed_logins (double)	logged_in (c
1	12345	0	0						
2	6548	8	15						
3	0	1	8						
4	7980	4	156						
5	555684	6	7						
6									

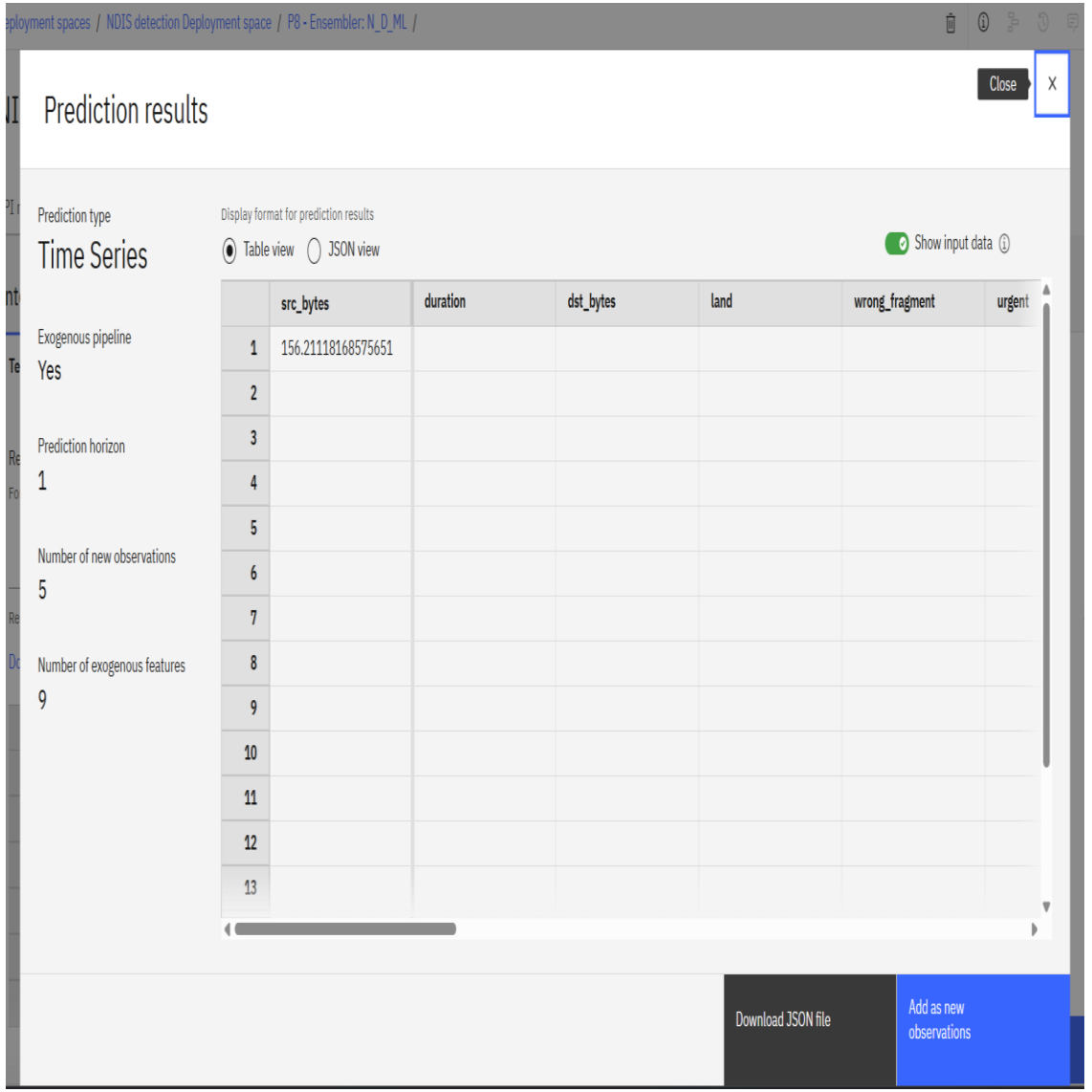
Predict

RESULT

- Present the results of the machine learning model in terms of its accuracy and effectiveness in detecting intrusions. The Ensembler topped the leaderboard with:
- Validation SMAPE: 103.405
- Holdout SMAPE: 102.876
- Backtest SMAPE: 102.616
- Sample predictions (e.g., for src_bytes inputs) show values like 248.13356508900128, demonstrating distinction between normal and malicious traffic.


dataplatform.cloud.ibm.com/ml-runtime/deployments/249837f3-a0b9-4bd8-bff7-bb1edf771aba/test?space_id=d5ff021e-c2c0-430d-8420-190e059247e3&context=cpdaas&flush=true
IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Ambarapu Humaun Basha'...
Dallas
AH
Deployment spaces / NDIS detection Deployment space / P8 - Ensembler: N_D_ML /
Prediction results
Prediction type
Time Series
Exogenous pipeline
Yes
Prediction horizon
1
Number of new observations
1
Number of exogenous features
9
Display format for prediction results
Table view
JSON view
Show input data
JSON view
[
{
"fields": [
"prediction"
],
"values": [
[
[
248.13356508900128
]
]
]
}
]
Download JSON file
Add as new observations





edunet foundation

CONCLUSION

- The Effectiveness of the proposed NIDS, The Ensembler model provides robust detection of cyber-attacks, leveraging IBM Cloud for accuracy. Challenges included data imbalance and real-time processing, addressed via AutoAI. This emphasizes the importance of ML in securing networks against threats.

FUTURE SCOPE

This include incorporating additional data sources, optimizing the algorithm for better performance, and expanding to cover IoT or mobile networks. Consider integration of emerging technologies such as edge computing or advanced deep learning techniques.

REFERENCES

- This include the Kaggle dataset, IBM AutoAI documentation, and papers on network intrusion detection, machine learning algorithms, and best practices in data preprocessing and model evaluation.
- Dataset: https://www.kaggle.com/datasets/sampadab17/networkintrusion-detection

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Ambarapu Humaun Basha

Has successfully satisfied the requirements for:

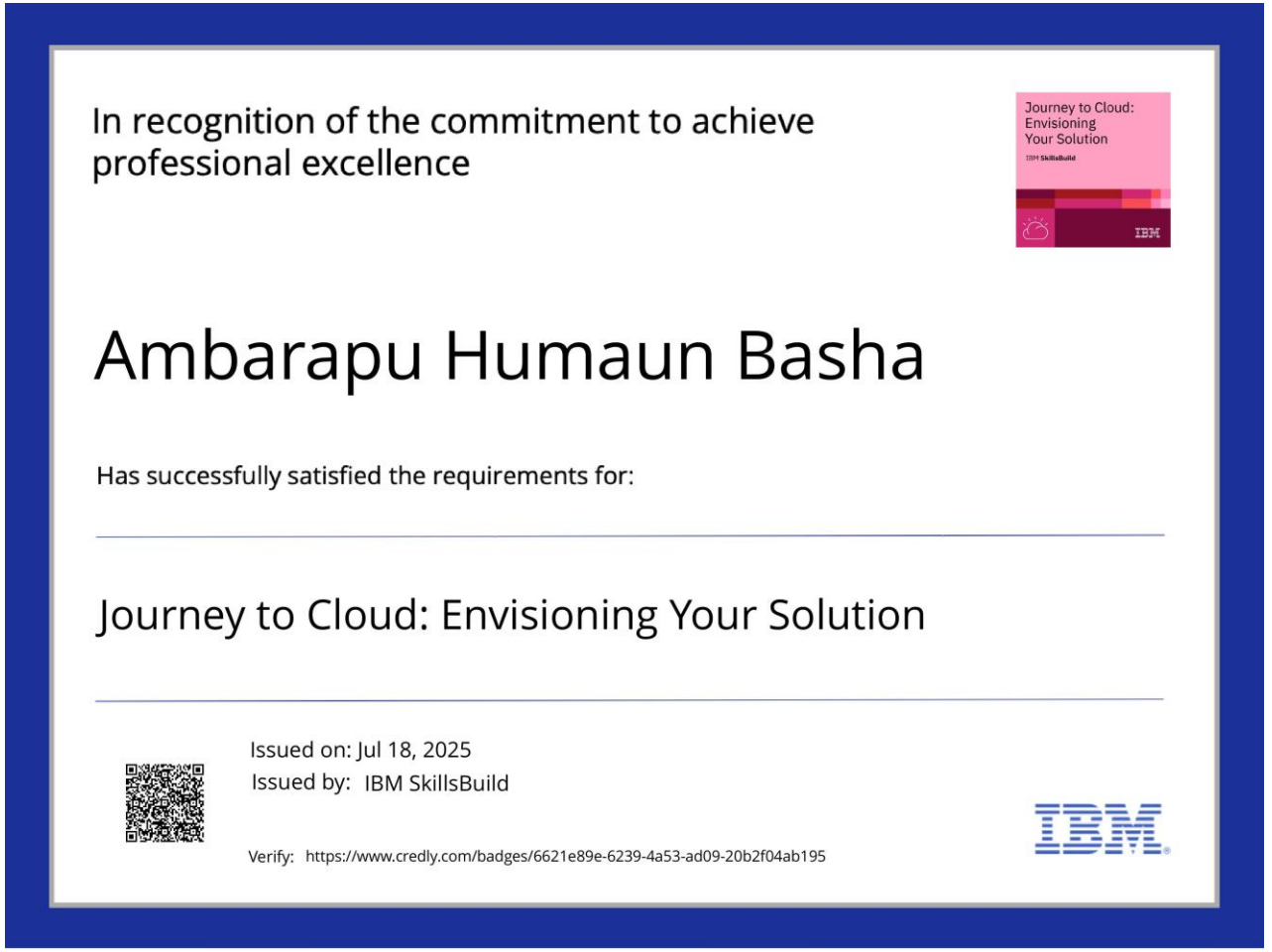
Getting Started with Artificial Intelligence

Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/9efb7cdf-dc0e-439b-97bf-78f25d873efb

IBM

IBM CERTIFICATIONS



IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Ambarapu Humaun Basha

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 18 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU

edunet foundation