CAPSTONE PROJECT

NETWORK INTRUSION DETECTION SYSTEM

Presented By:
1. Pavithra V-City Engineering College -CSE

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

In today’s digital age, computer networks are highly vulnerable to cyber-attacks such as Denial of Service (DoS), Remote-to-Local (R2L), User-to-Root (U2R), and Probe attacks. Traditional intrusion detection methods often fail to detect unknown or sophisticated threats in real time. There is a need for an intelligent and automated system to identify and classify malicious network activity from normal behavior, ensuring timely and accurate detection.

PROPOSED SOLUTION

The proposed system aims to address the challenge of detecting and classifying malicious network activity using machine learning. The goal is to improve network security by automatically identifying potential cyber-attacks such as DoS, R2L, U2R, and Probe attacks. The system leverages supervised learning techniques and cloud-based tools to build a scalable and intelligent intrusion detection solution. The solution will consist of the following components:

- **Data Collection:**
  - Gather labeled network traffic data from publicly available sources such as the **KDD Cup 1999 dataset** (available on Kaggle).
  - The dataset includes features such as protocol type, service, flag, duration, source bytes, and destination bytes, along with labels identifying normal and attack types.
  - Store and manage the dataset using **IBM Cloud Object Storage**.

PROPOSED SOLUTION

- **Data Preprocessing:**
  - Clean the dataset by handling missing values, removing duplicates, and correcting inconsistencies.
  - Encode categorical features (e.g., protocol type, service, flag) into numerical form.
  - Normalize the feature values to ensure uniform scale for model training.
  - Simplify attack labels into major categories (e.g., DoS, Probe, R2L, U2R, Normal) for better classification accuracy.
- **Machine Learning Algorithm:**
  - Use **IBM Watson Studio's AutoAI** to automatically select and train the best model.
  - AutoAI evaluates multiple algorithms including **Random Forest**, **Logistic Regression**, and **Gradient Boosted Trees** based on performance metrics.
  - The system selects the most accurate model for classifying normal and attack traffic.
  - IBM Watson Machine Learning is used to deploy the final model as a cloud-based API.

PROPOSED SOLUTION

- **Deployment:**
  - Deploy the trained model on **IBM Watson Machine Learning** with an **online REST API endpoint**.
  - Use the endpoint to allow real-time or batch predictions for new network traffic data.
  - The model can be integrated into a security dashboard or backend system for real-time monitoring.
- **Evaluation:**

  Assess the model using metrics such as:
  - Accuracy
  - Precision
  - Recall
  - Use the Watson Studio AutoAI leaderboard and model evaluation section to compare models and select the best-performing pipeline.

PROPOSED SOLUTION

- **Result:**
  - The system successfully classifies different types of intrusions with high accuracy.
  - Visual results such as a **confusion matrix** and **classification report** are generated automatically.
  - Users can test the deployed model with sample data through the **Test tab** or via the API endpoint.

SYSTEM APPROACH

Technology Used:

- **Cloud Platform**: IBM Cloud Lite (Free Student Tier)
- **Data Source**: Kaggle – KDD Cup 1999 Intrusion Detection Dataset
- **Development Environment**: IBM Watson Studio (AutoAI)
- **Model Deployment**: IBM Watson Machine Learning (Online Deployment)
- **Data Storage**: IBM Cloud Object Storage
- **Evaluation**: AutoAI metrics (Accuracy, Precision, Recall, Confusion Matrix)

ALGORITHM & DEPLOYMENT

- **Algorithm Used:**

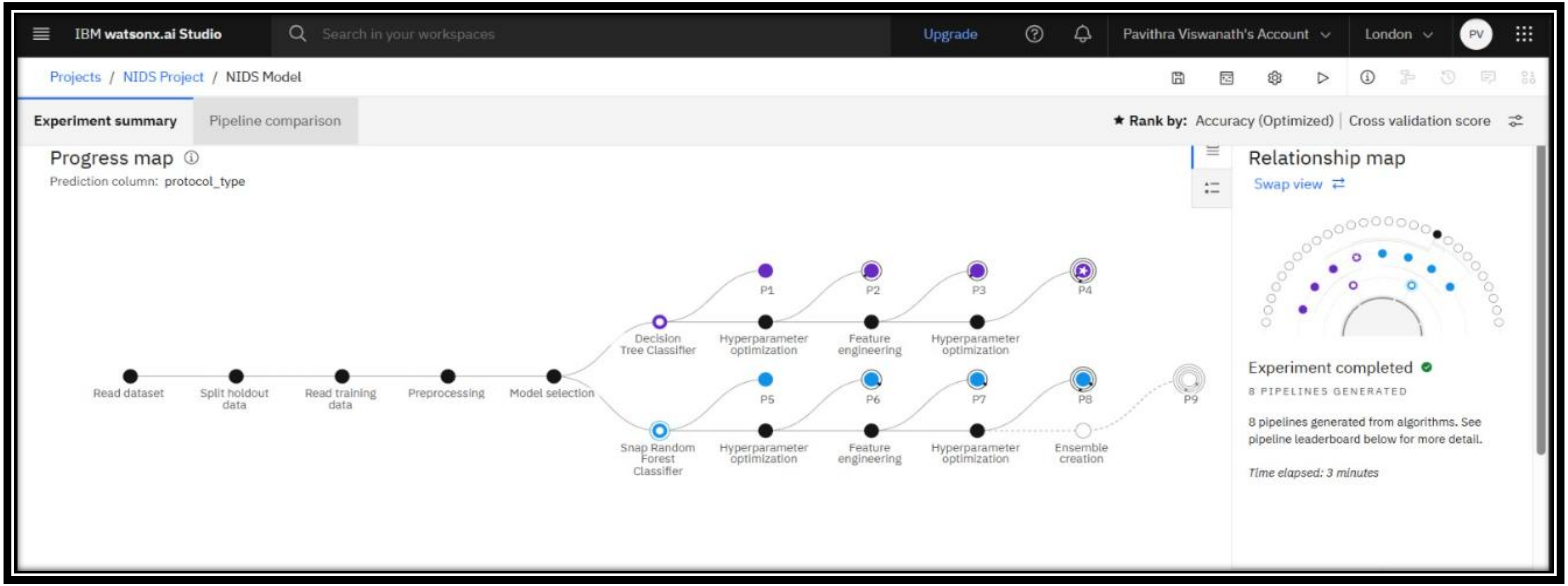
  AutoAI in Watson Studio selects from:

  - Random Forest
  - Decision Tree Classifier
  - Logistic Regression
  - Gradient Boosting
    (AutoAI automatically chooses the best based on performance)

- **Model Training:**
  - Data uploaded and preprocessed automatically.
  - Best model selected via Leaderboard based on accuracy and F1 score.

- **Deployment:**
  - The trained model is deployed as an online endpoint in IBM Watson Machine Learning.
  - Allows API-based testing of network traffic features for prediction.

RESULT



★ Rank by: Accuracy (Optimized) | Cross validation score

Prediction column: protocol_type

Experiment completed

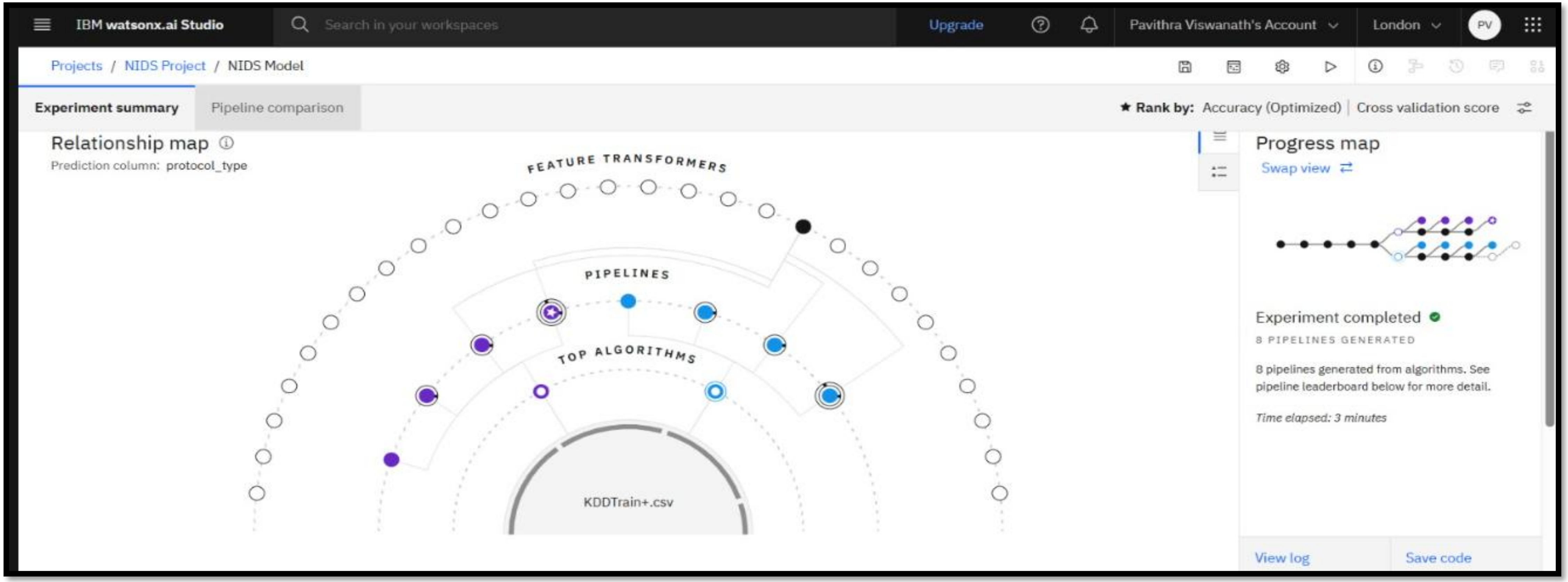
8 PIPELINES GENERATED

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

Time elapsed: 3 minutes

The AutoAI progress map shows the automated machine learning workflow for the NIDS project using IBM Watsonx.ai. It generated 8 optimized pipelines using Decision Tree and Snap Random Forest classifiers. The pipelines were evaluated and ranked based on accuracy for model deployment.

RESULT



Experiment completed

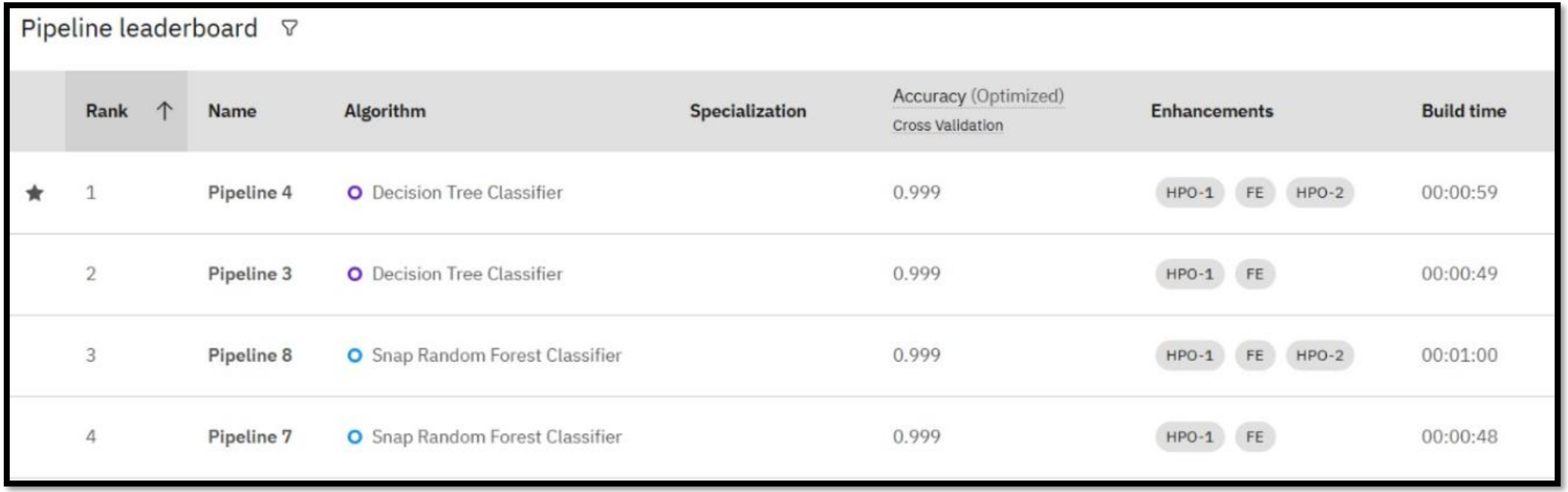
8 PIPELINES GENERATED

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

Time elapsed: 3 minutes

The relationship map shows how AutoAI generated 8 ML pipelines using top algorithms like Decision Tree and Random Forest on the KDDTrain+.csv dataset. It visualizes the use of different feature transformers and optimization steps. The pipelines were ranked based on accuracy to select the best model.

RESULT

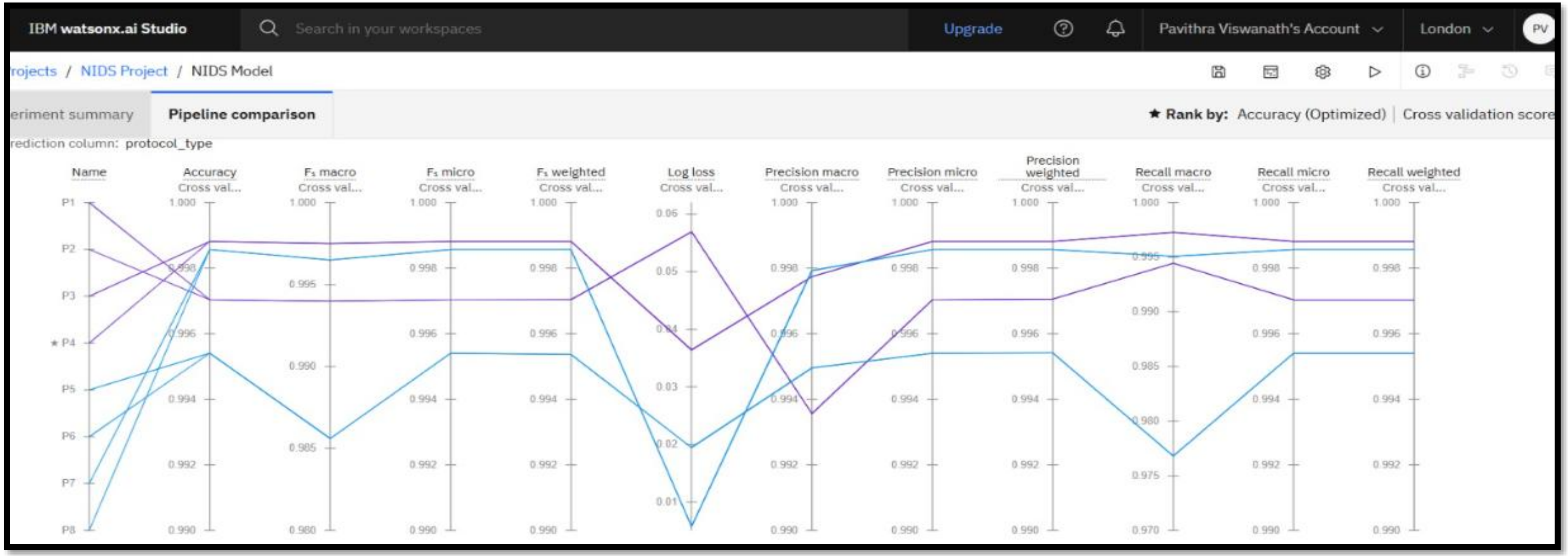


Pipeline leaderboard

	Rank ↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 4	○ Decision Tree Classifier		0.999	HPO-1 FE HPO-2	00:00:59
	2	Pipeline 3	○ Decision Tree Classifier		0.999	HPO-1 FE	00:00:49
	3	Pipeline 8	○ Snap Random Forest Classifier		0.999	HPO-1 FE HPO-2	00:01:00
	4	Pipeline 7	○ Snap Random Forest Classifier		0.999	HPO-1 FE	00:00:48

The pipeline leaderboard displays the top-performing models ranked by optimized accuracy. Pipeline 4, using a Decision Tree Classifier, achieved the highest accuracy of 0.999. All top pipelines include enhancements like hyperparameter optimization (HPO) and feature engineering (FE).

RESULT

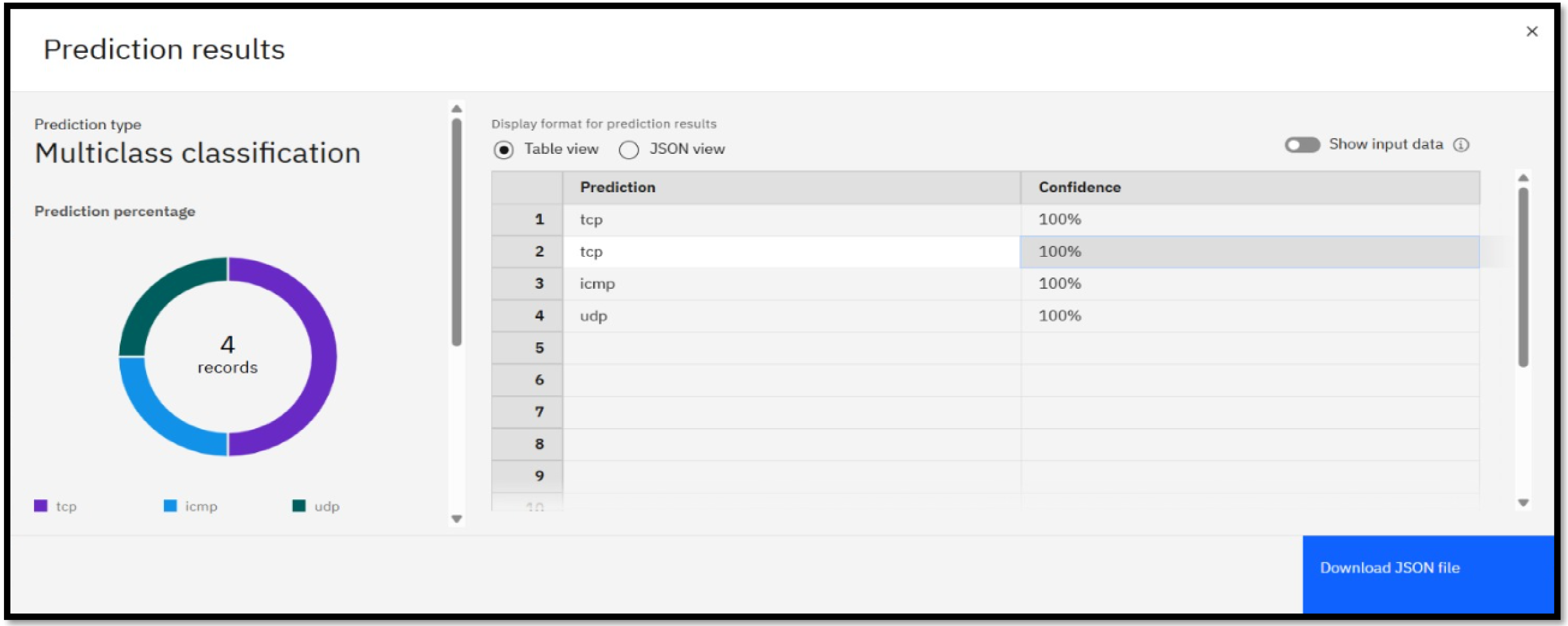


★ Rank by: Accuracy (Optimized) | Cross validation score

rediction column: protocol_type

The pipeline comparison graph displays key evaluation metrics for all models. Pipeline 4 shows top performance across accuracy, precision, recall, and F1 scores with low log loss, making it the best choice. The visual comparison confirms its consistent superiority over other pipelines.

RESULT



Prediction results

Prediction type

Multiclass classification

Prediction percentage

Display format for prediction results

● Table view ○ JSON view

Show input data

	Prediction	Confidence
1	tcp	100%
2	tcp	100%
3	icmp	100%
4	udp	100%
5		
6		
7		
8		
9		

The prediction results display a multiclass classification output for 4 network traffic records. Each record is accurately classified as either tcp, icmp, or udp with 100% confidence. The results confirm the model's strong predictive performance on protocol type.

CONCLUSION

- The proposed NIDS solution successfully classifies network traffic into normal and attack types with high accuracy using machine learning techniques. It demonstrates the effectiveness of AI-driven methods in detecting various types of cyber threats such as DoS, Probe, R2L, and U2R attacks.
- By leveraging **IBM Watson Studio's AutoAI**, the development process becomes beginner-friendly, eliminating the need for manual algorithm selection and parameter tuning. This enables faster model development and testing.
- The use of **IBM Cloud Lite services**, such as **Cloud Object Storage** and **Watson Machine Learning**, ensures scalability, reliability, and cost-effectiveness. The model can be easily deployed as an **online REST API**, making it accessible for integration with real-time monitoring systems or security dashboards.
- This system is a significant step toward automating network security and reducing manual effort in identifying potential threats. It aids network administrators and cybersecurity professionals by providing timely and accurate intrusion alerts, thereby enhancing the overall security posture.
- With high evaluation metrics (like precision, recall, and F1-score), the deployed model proves to be a reliable tool for intrusion detection in small to medium-sized networks.

FUTURE SCOPE

- **Extend to Real-Time Detection**

  Integrate the system with routers or firewalls to capture and analyze **live network traffic** using packet sniffing tools or streaming APIs.

- **Adopt Deep Learning Techniques**

  Implement models like **CNN**, **LSTM**, or **Autoencoders** to improve the detection of complex and evolving threats by automatically learning deeper traffic patterns.

- **Enhance Attack Classification Granularity**

  Improve the classification system to detect **specific attack types** (e.g., DDoS, SQL Injection, Brute Force) rather than broad categories.

- **Integrate with SIEM Tools**

  Connect the system with **Security Information and Event Management (SIEM)** platforms such as Splunk or IBM QRadar for centralized threat monitoring and analysis.

REFERENCES

[1] **KDD Cup 1999 Dataset-** https://www.kaggle.com/datasets/sampadab17/networkintrusion-detection

[2] **IBM Cloud Platform-** Cloud platform used for deploying machine learning models and hosting storage.

[3] **IBM Watson Studio Documentation-** https://www.ibm.com/docs/en/watson-studio

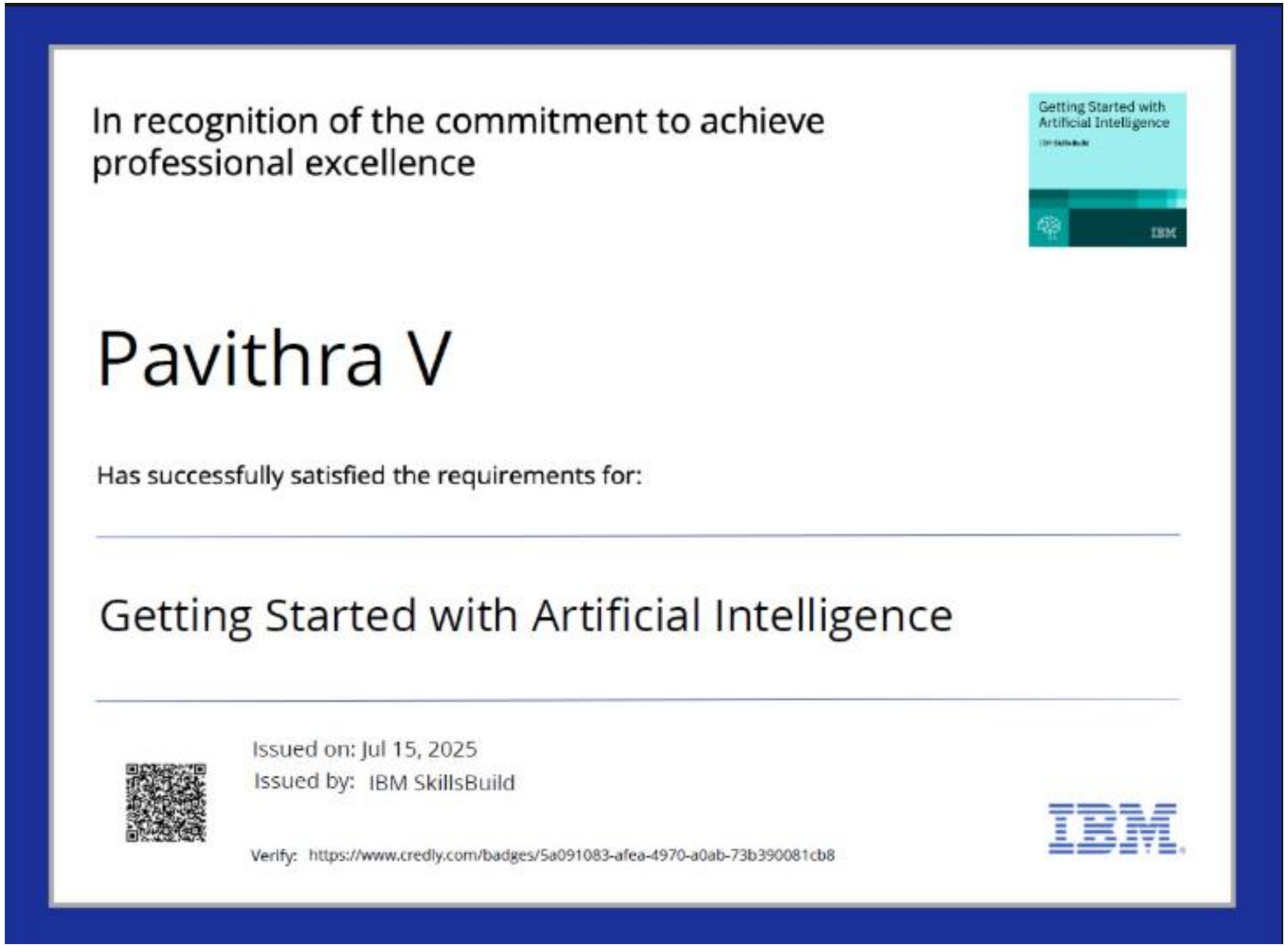
[4] **IBM Watson Machine Learning Documentation**- https://www.ibm.com/docs/en/watson-machine-learning

[5] **IBM Watsonx.ai (Granite LLM)-** https://www.ibm.com/products/watsonx

[6] **IBM Cloud Object Storage Documentation**- https://cloud.ibm.com/docs/cloud-object-storage

IBM CERTIFICATIONS

Getting Started With Artificial Intelligence



Getting Started with Artificial Intelligence

IBM SkillsBuild

IBM

In recognition of the commitment to achieve professional excellence

Pavithra V

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025

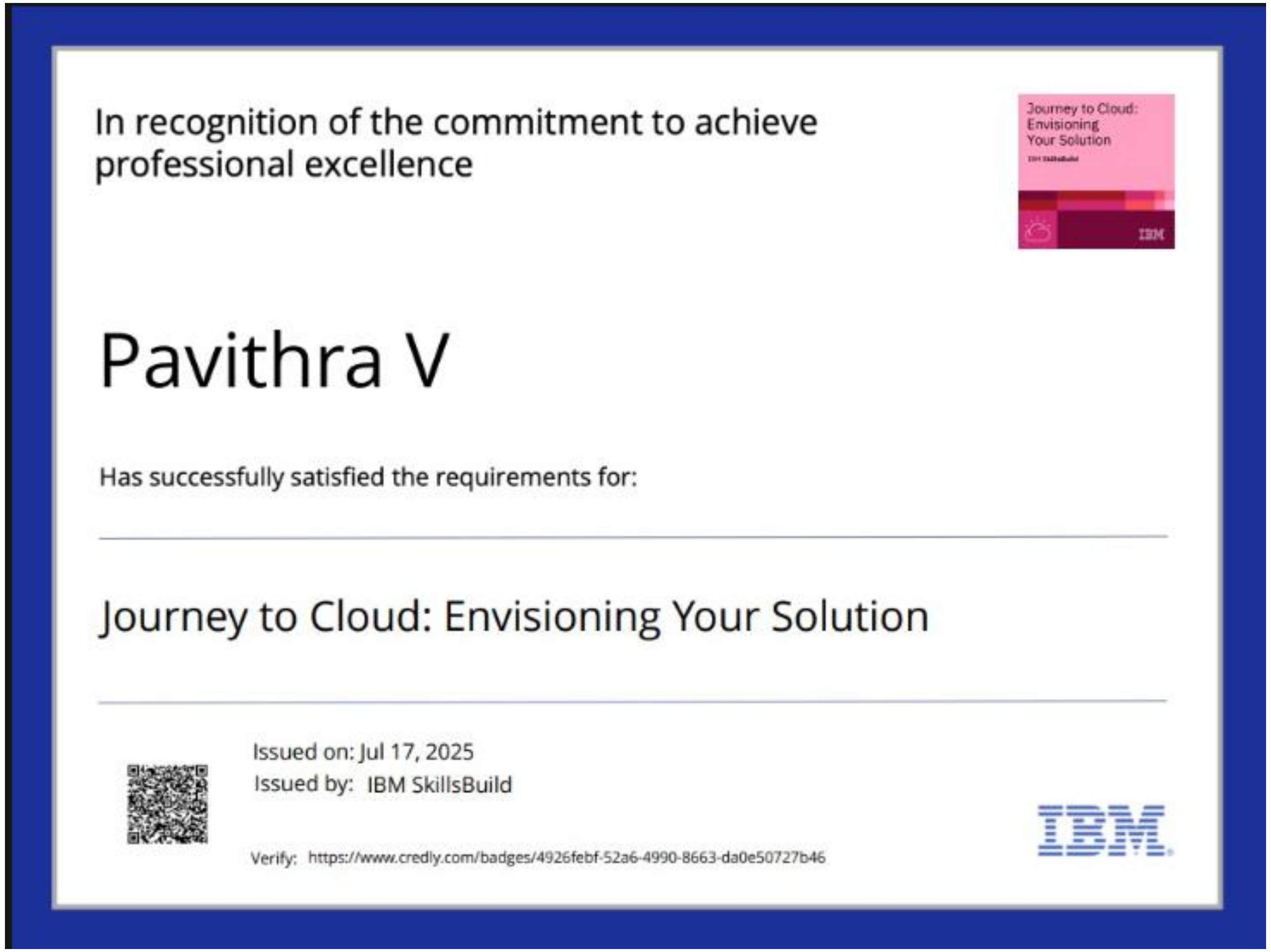
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/5a091083-afea-4970-a0ab-73b390081cb8

IBM

IBM CERTIFICATIONS

Journey to Cloud: Envisioning Your Solution


Journey to Cloud:
Envisioning
Your Solution
IBM SkillsBuild
IBM

In recognition of the commitment to achieve professional excellence

Pavithra V

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

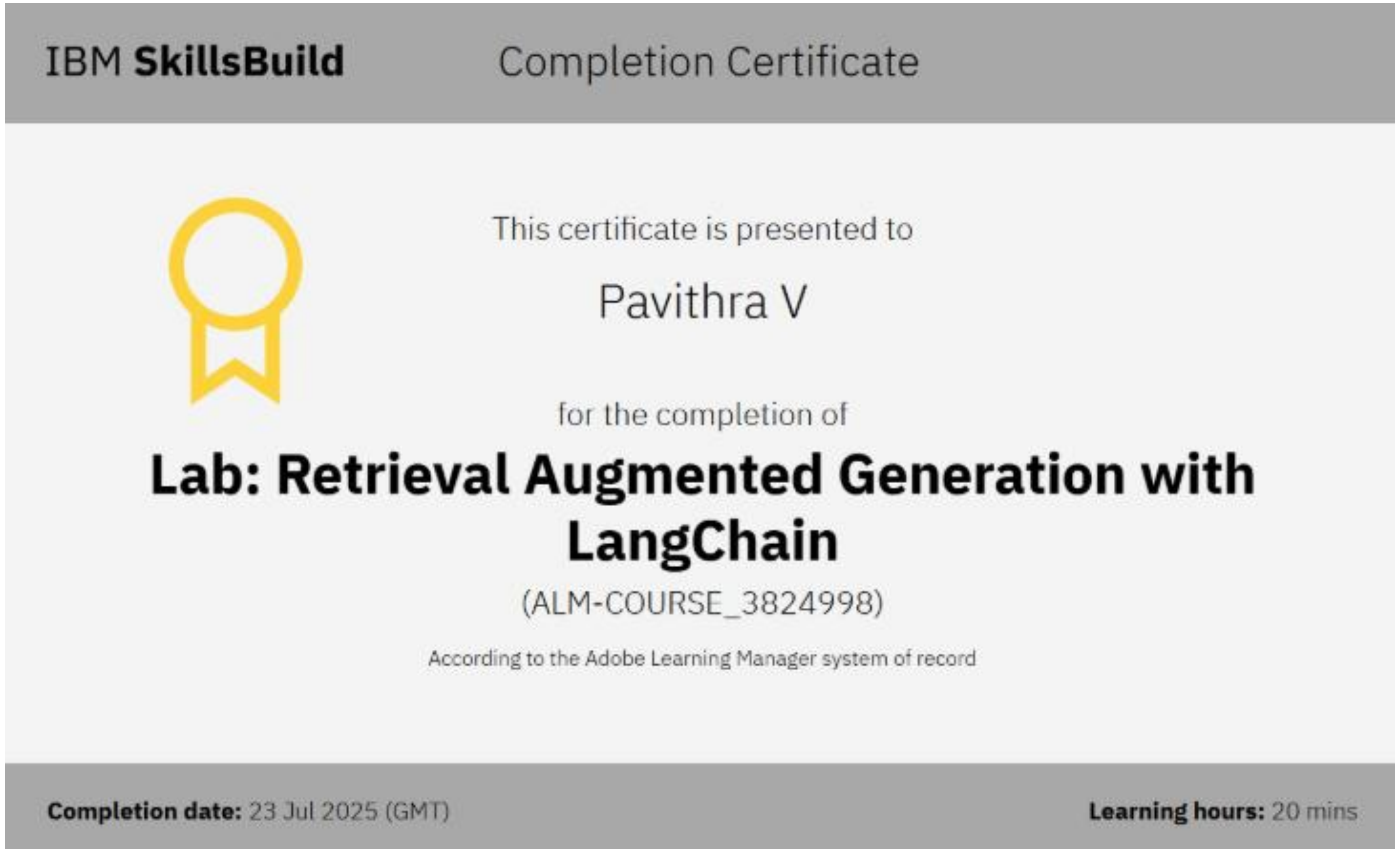
Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/4926febf-52a6-4990-8663-da0e50727b46

IBM

IBM CERTIFICATIONS

RAG Lab with LangChain



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Pavithra V

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