

# CAPSTONE PROJECT

## NETWORK INTRUSION DETECTION

### **Presented By:**

**Student Name :- Ritika Singh**

**College Name :- SANDIP UNIVERSITY**

**Department :- BTech(CSE)**

**AICTE Student ID: STU6857ab2c1b8881750575916**

**Internship ID: INTERNSHIP\_1748937226683eaa0a58abc**

**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

The screenshots illustrate the workflow for setting up and using IBM watsonx.ai Studio. The process begins with navigating to the IBM Cloud console, where the 'watsonx.ai Studio-z8' resource is identified. A 'Launch in' button is available to initiate the setup. The next step involves the 'Welcome, Ritika!' dashboard, which provides a 'Get started' modal to guide the user through the initial configuration. This modal offers options to either open a sample project or create a new project. The 'Create a project' dialog is shown, allowing users to define project details such as Name, Description, Tags, and Storage. Finally, the 'Welcome, Ritika!' dashboard is shown again, highlighting the 'Get started' modal with options to provision watsonx.ai Studio or Runtime from the service catalog.

# SCREENSHOTS OF WORKFLOW

The screenshots illustrate the workflow in IBM watsonx.ai Studio for a project named "Network intrusion detection".

**Screenshot 1: Project Overview**  
The "Start working" section provides quick actions: "Add users as collaborators", "Add data to work with", "Work with data and models in Python or R notebooks", and "Build machine learning models automatically". The "Assets" section shows a bar chart with 0 CUH. The "Resource usage" section shows 0 CUH. The "Your documentation" section includes a link to "Open Documentation editor". The "Project history" section shows the project was created at 2:15 PM.

**Screenshot 2: Project Management**  
The "Manage" tab shows project details for "ad110304-602f-4766-ab98-2a1b27f1ccaf". The "Controls" section includes options to "Enable folders", "Switch platform", and "Grant access". The "Danger zone" section includes options to "Leave project" or "Delete project".

**Screenshot 3: Associate service**  
The "Associate service" dialog shows a table of available services:

Name	Type	Plan	Location	Status	Group
watsonx.ai Runtime-pw	watsonx.ai Runtime	Lite	Sydney	Not associated	Default

**Screenshot 4: Build machine learning models automatically**  
The "Build machine learning models automatically" dialog shows the "Define details" and "Define configuration" sections. The "Define details" section includes fields for "Name", "Description (optional)", and "Tags (optional)". The "Define configuration" section includes fields for "watsonx.ai Runtime service Instance" and "Environment definition".

**Screenshot 5: Build machine learning models automatically (continued)**  
The "Define configuration" section shows the "Environment definition" as "Large: 8 CPU and 32 GB RAM". It also notes that this environment definition consumes 20 capacity units per hour for training.



# SCREENSHOTS OF WORKFLOW

This screenshot shows the 'Configure AutoAI experiment' interface for a 'Network Intrusion system'. The 'Add data source' section on the left shows 'Train\_data.csv' (Size: 2.74 MB, Columns: 42) with 'Browse' and 'Select from project' buttons. The 'Configure details' section on the right includes a 'Create a time series analysis?' toggle (set to 'No'), a 'What do you want to predict?' section with 'Prediction column' set to 'class', and a 'Prediction column class' dropdown set to 'Binary Classification'. The 'Experiment settings' section at the bottom shows 'Prediction type' as 'Binary Classification', 'Positive class' as 'Normal', and 'Optimized for' as 'Accuracy & run time'. A 'Run experiment' button is at the bottom right.

This screenshot shows the 'Configure AutoAI experiment' interface for a 'Network Intrusion system'. The 'Add data source' section on the left shows 'Train\_data.csv' (Size: 2.74 MB, Columns: 42) with 'Browse' and 'Select from project' buttons. The 'Configure details' section on the right includes a 'Create a time series analysis?' toggle (set to 'No'), a 'What do you want to predict?' section with 'Prediction column' set to 'class', and a 'Prediction column class' dropdown set to 'Binary Classification'. The 'Experiment settings' section at the bottom shows 'Prediction type' as 'Binary Classification', 'Positive class' as 'Normal', and 'Optimized for' as 'Accuracy & run time'. A 'Run experiment' button is at the bottom right.

This screenshot shows the 'Experiment summary' and 'Pipeline comparison' tabs. The 'Experiment summary' tab displays a flowchart of the experiment process: Read dataset -> Split holdout data -> Read training data -> Preprocessing -> Model selection -> Snap Decision Tree Classifier -> Hyperparameter optimization -> Feature engineering -> Hyperparameter optimization -> Hyperparameter optimization -> Hyperparameter optimization. The 'Pipeline comparison' tab shows a table of generated pipelines.

Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
1	Pipeline 2	Snap Decision Tree Classifier	0.995	HPO-1	00:01:00
2	Pipeline 1	Snap Decision Tree Classifier	0.995	None	00:00:02
3	Pipeline 6	Decision Tree Classifier	0.994	HPO-1	00:00:08
4	Pipeline 5	Decision Tree Classifier	0.994	None	00:00:04

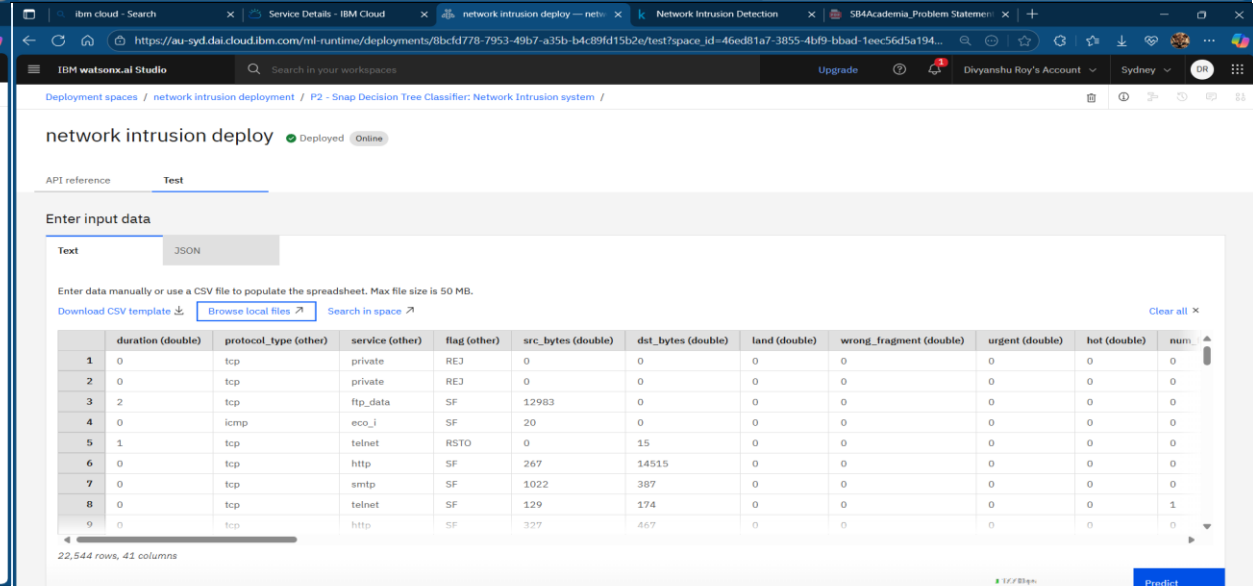
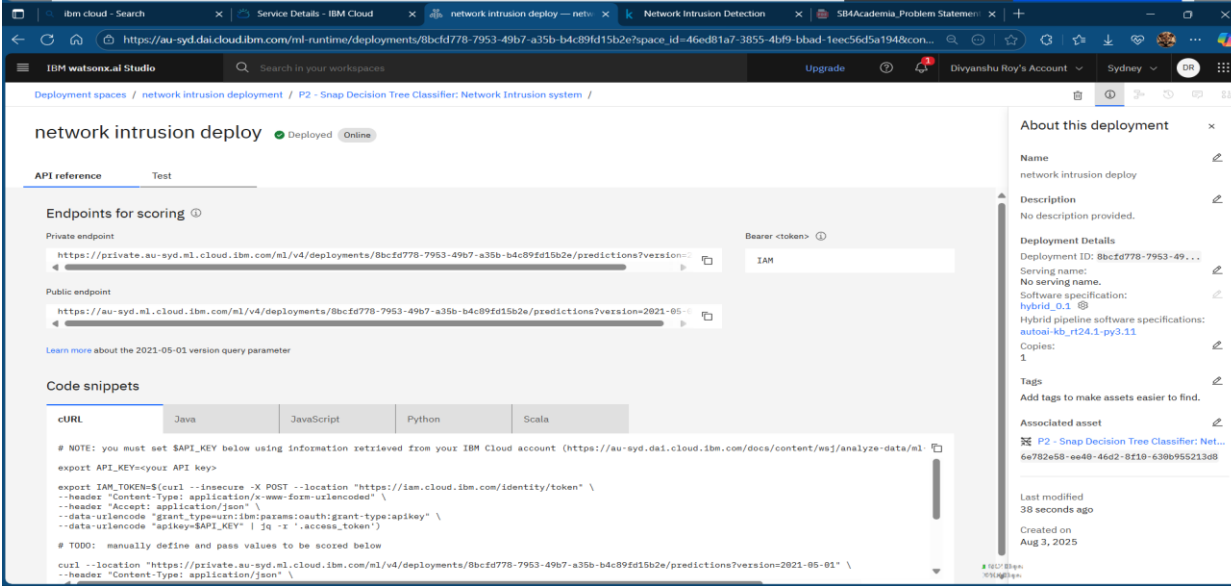
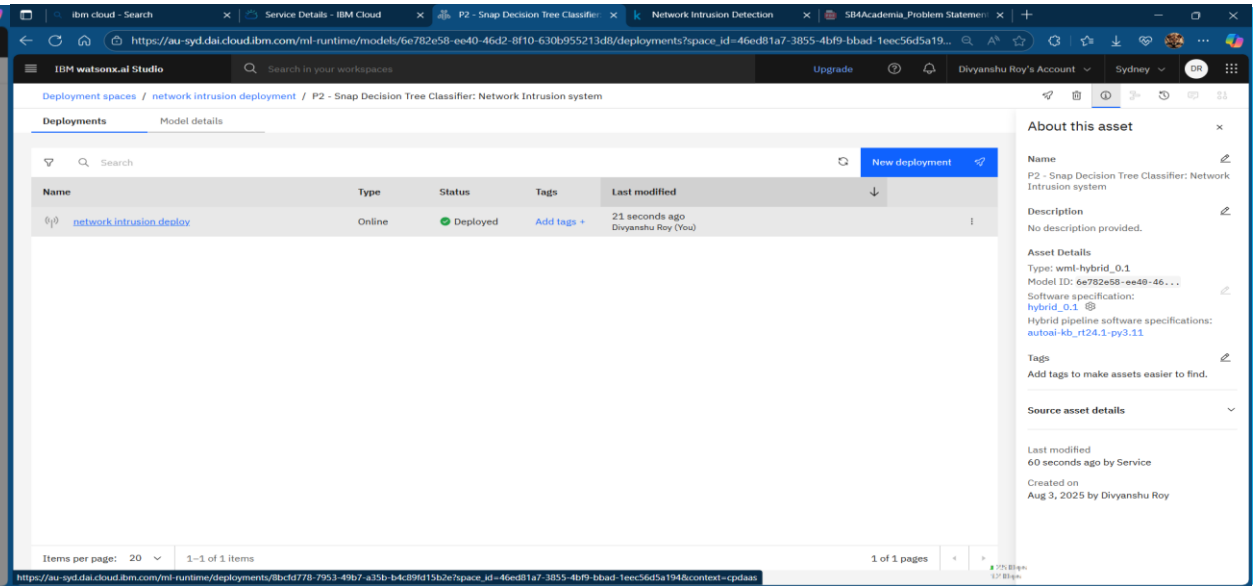
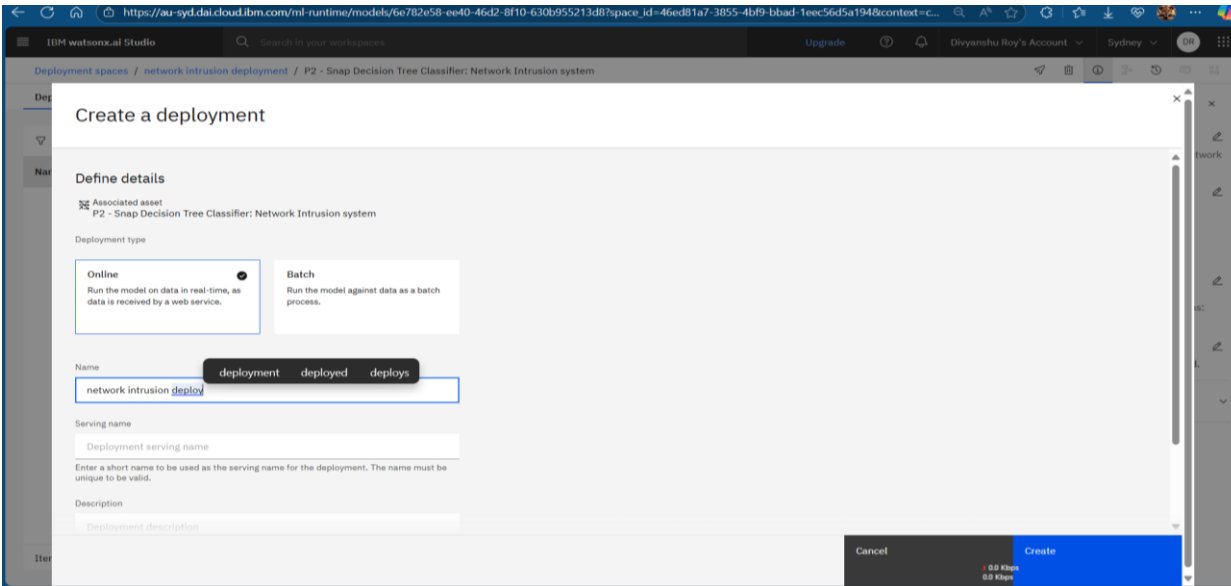
This screenshot shows the 'Save as' dialog. The 'Select asset type' section has 'Model' selected. The 'Define details' section shows 'Name' as 'P2 - Snap Decision Tree Classifier: Network Intrusion system' and 'Description (optional)' as 'Model description'. The 'Tags' section has 'Add tags to make assets easier to find.' and 'Add a tag' button. The 'Create' button is at the bottom right.

# SCREENSHOTS OF WORKFLOW

The screenshots illustrate the process of promoting an asset to a deployment space in IBM Watson AI Studio:

- Top Left:** Shows the 'Input (1)' table with columns 'Column' and 'Type'. The table lists various metrics like 'count', 'diff\_srv\_rate', 'dst\_bytes', etc., all of type 'double'. A sidebar on the right shows asset details for 'P2 - Snap Decision Tree Classifier: Network Intrusion system', including its description, asset details (Type: wml-hybrid\_0.1, Model ID: 81c09951-aad2-4d91-a0dd-d5002fd92b8e), software specification (hybrid\_0.1), and tags.
- Top Right:** Shows the 'Promote to space' dialog box. It prompts the user to 'Create a deployment space' and provides instructions on how to use the space. A 'Define details' section is visible, showing the name 'network intrusion deployment' and a description 'The space "network intrusion deployment" is being created.'
- Bottom Left:** Shows the 'Promote to space' dialog box after the promotion is completed. A green success message states: '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. Timestamp 10:58:58 AM'. Below the message, a table shows the 'Selected assets (1)' with columns 'Name', 'Format', 'Version', and 'Status'. The asset 'P2 - Snap Decision Tree Classifier: Network Intrusion system' is listed with format 'Model', version 'Current', and status 'Promoted'.
- Bottom Right:** Shows the 'network intrusion deployment' page. It displays the 'Assets' tab, showing a list of assets. The asset 'P2 - Snap Decision Tree Classifier: Network Intrusion system' is listed with details like 'Machine learning model from AutoAI' and 'Last modified 8 seconds ago'.

# SCREENSHOTS OF WORKFLOW



# 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'.

The screenshot displays the IBM watsonx.ai Studio interface. The top navigation bar includes the IBM logo, 'watsonx.ai Studio', a search bar, and user account information for 'Ritika Singh's Account'. The breadcrumb trail indicates the current location: 'Deployment spaces / network intrusion deployment space / P2 - Snap Decision Tree Classifier: Network intrusion detection system /'. The main content area is titled 'Prediction results' and features a 'Close' button. Below the title, there are options to 'Display format for prediction results' (Table view is selected) and a toggle for 'Show input data'. The table below shows 15 rows of prediction results, each with an index, a prediction ('normal' or 'anomaly'), and a probability array (e.g., [1,0] for anomaly, [0,1] for normal). A 'Download JSON file' button is located at the bottom right, with a file size indicator showing '0.5 Kbps' and '1.7 Kbps'.

	prediction	probability
1	anomaly	[1,0]
2	anomaly	[1,0]
3	normal	[0,1]
4	anomaly	[1,0]
5	normal	[0,1]
6	normal	[0,1]
7	normal	[0,1]
8	normal	[0,1]
9	normal	[0,1]
10	anomaly	[1,0]
11	anomaly	[1,0]
12	normal	[0,1]
13	anomaly	[1,0]
14	anomaly	[1,0]
15	normal	[0,1]

# 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



## Ritika Singh

Has successfully satisfied the requirements for:

### Getting Started with Artificial Intelligence



Issued on: Jul 18, 2025  
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/f81d8579-8238-41b5-b3f0-1980023c0826>





# IBM CERTIFICATIONS

In recognition of the commitment to achieve  
professional excellence



## Ritika Singh

Has successfully satisfied the requirements for:

---

### Journey to Cloud: Envisioning Your Solution

---



Issued on: Jul 18, 2025

Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/ade7b062-f1d7-432e-916c-279b87e9e997>



# IBM CERTIFICATIONS

**IBM SkillsBuild** Completion Certificate



This certificate is presented to

**Ritika Singh**

for the completion of

**Lab: Retrieval Augmented Generation with  
LangChain**

(ALM-COURSE\_3824998)

According to the Adobe Learning Manager system of record

**Completion date:** 24 Jul 2025 (GMT) **Learning hours:** 20 mins



**THANK YOU**