
CAPSTONE PROJECT

NETWORK INTRUSION DETECTION USING MACHINE LEARNING

Presented By:

1. Rudraksh Dattaram Chorge - VPM's BN Bandodkar College Of Science – BSC CS

OUTLINE

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

PROBLEM STATEMENT

Modern network infrastructures are increasingly exposed to a wide range of cyber-attacks such as Denial of Service (DoS), Probing, Remote-to-Local (R2L), and User-to-Root (U2R). These attacks can compromise data integrity, service availability, and system confidentiality. The key challenge is to accurately distinguish between normal and malicious network traffic using large volumes of connection records and behavioral patterns, without relying on manual monitoring. Failure to do so can result in undetected intrusions and potential breaches across critical systems.

PROPOSED SOLUTION

- The proposed system aims to address the critical challenge of automatically detecting and classifying malicious network traffic in order to strengthen cybersecurity defenses. Leveraging machine learning and data-driven techniques, the solution is designed to analyze historical connection data and predict whether a network activity is normal or an intrusion attempt.
- The system includes the following components:
- **Data Collection:**
Utilized the NSL-KDD dataset, which is a refined version of the KDD Cup 1999 dataset, containing labeled records of network traffic across various attack types. The dataset includes 42 input features capturing characteristics such as protocol type, service, connection duration, byte counts, and error rates.
- **Data Preprocessing:**
Cleaned and transformed the dataset to handle categorical variables using label encoding and normalized numerical values using feature scaling. Verified data integrity and ensured all records were consistently formatted for training.
- **Model Training & Algorithm Selection:**
Employed **IBM Watson AutoAI**, which automatically tested multiple algorithms such as Random Forest, Gradient Boosted Trees, and Logistic Regression. The best-performing model was selected based on accuracy and evaluation metrics. The system supports multiclass classification across 23 types of attacks and normal traffic.
- **Deployment:**
The trained model was deployed via **IBM CLOUD**
- **Evaluation:**
The model's performance was assessed using metrics like **accuracy**, **precision**, **recall**, and **confusion matrix**. Real input samples were tested through the deployment interface, and output responses were captured and documented.

SYSTEM APPROACH

- **Tools & Platforms:**

- IBM Cloud (Mandatory)
- IBM Watson Studio for model development and deployment.
- IBM Cloud Object Storage for dataset handling.

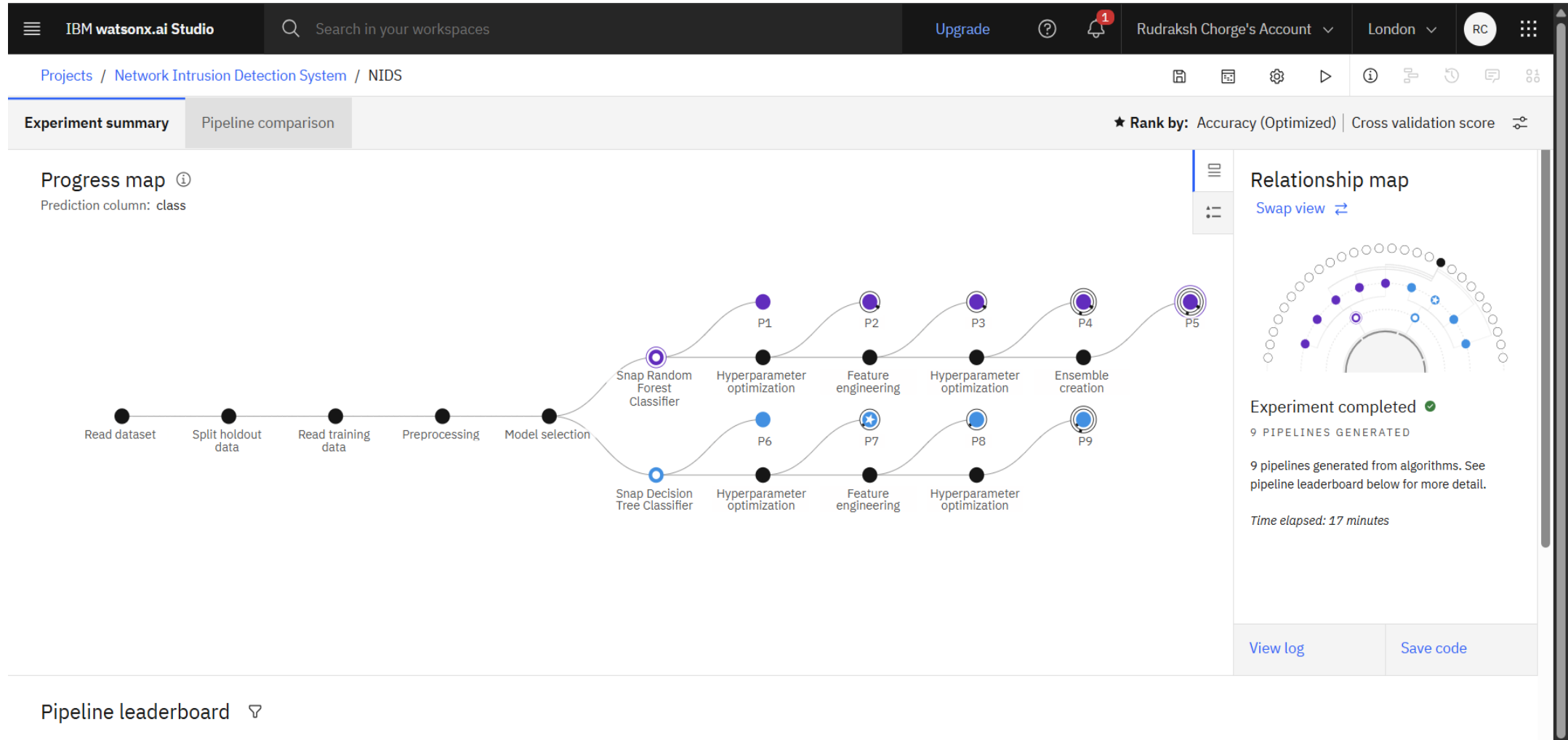
- **Dataset Used:**

- NSL-KDD Dataset (KDDTrain+) from Kaggle Dataset
<https://www.kaggle.com/datasets/hassan06/nsllkdd?select=KDDTrain%2B.txt>

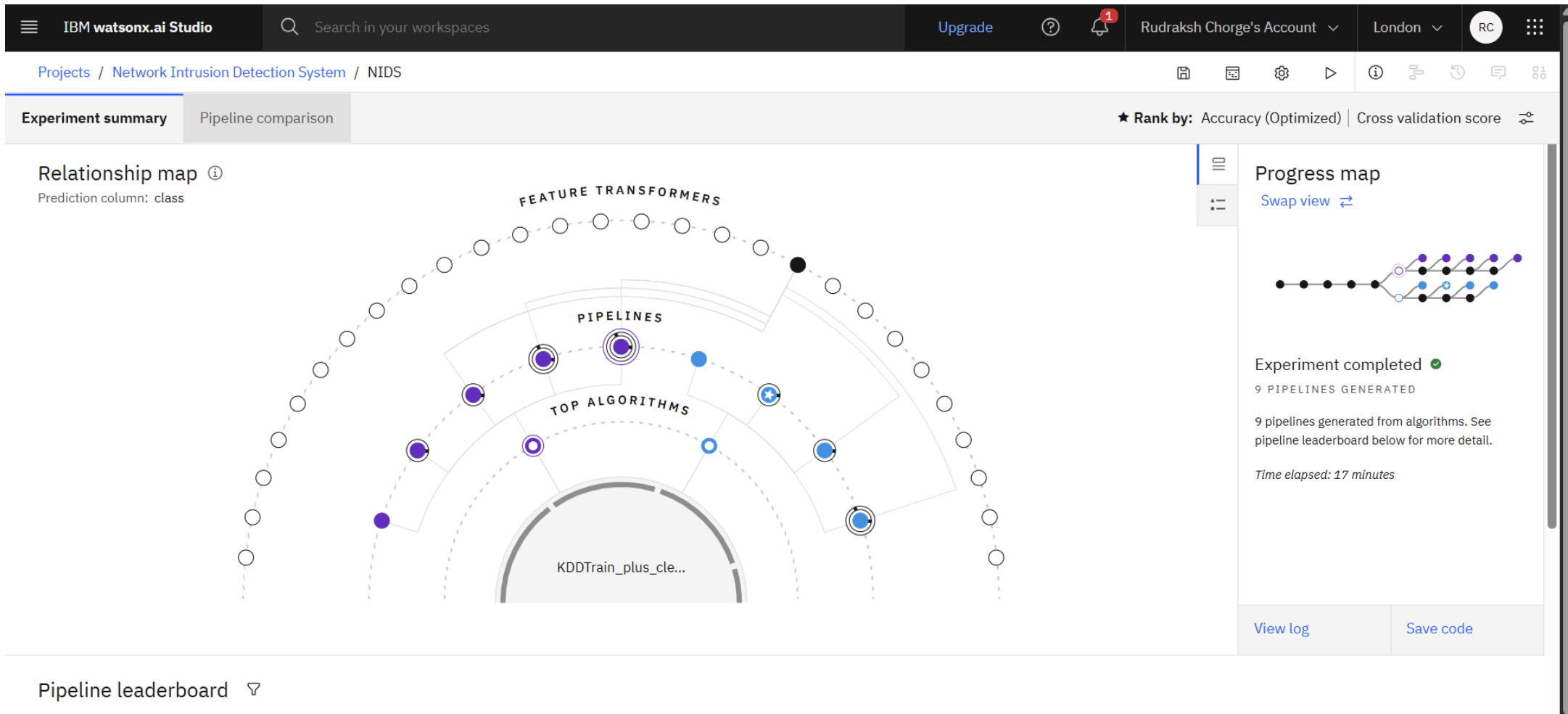
ALGORITHM & DEPLOYMENT

- **Algorithm Used:** Random Forest Classifier (multiclass classification)
- **Input Features:** Protocol type, service, duration, bytes sent/received, error rates, etc.
- **Output:** Label representing intrusion type or normal
- **Training Process:**
 - Categorical encoding + feature scaling
 - Train-test split (80-20)
- **Deployment:**
 - Trained model deployed on IBM Watson Studio with API endpoint for real time predictions.

RESULT



RESULT



RESULT

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

1

Rudraksh Chorge's Account

London

RC

Projects / Network Intrusion Detection System / NIDS

Experiment summary

Pipeline comparison

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

Snap Decision Tree Classifier

Hyperparameter optimization

Feature engineering

Hyperparameter optimization

Time elapsed: 17 minutes

View log

Save code

Pipeline leaderboard

| | Rank | Name | Algorithm | Specialization | Accuracy (Optimized) Cross Validation | Enhancements | Build time |
|---|------|------------|-------------------------------|----------------|--|----------------|------------|
| ★ | 1 | Pipeline 7 | Snap Decision Tree Classifier | | 0.997 | HPO-1 | 00:07:42 |
| | 2 | Pipeline 6 | Snap Decision Tree Classifier | | 0.997 | None | 00:07:18 |
| | 3 | Pipeline 9 | Snap Decision Tree Classifier | | 0.995 | HPO-1 FE HPO-2 | 00:00:24 |
| | 4 | Pipeline 8 | Snap Decision Tree Classifier | | 0.995 | HPO-1 FE | 00:09:53 |

RESULT

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

?

🔔

Rudraksh Chorge's Account

London

RC

Deployment spaces / NIDS / P7 - Snap Decision Tree Classifier: NIDS

Network Intrusion Detector ✔ 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](#)

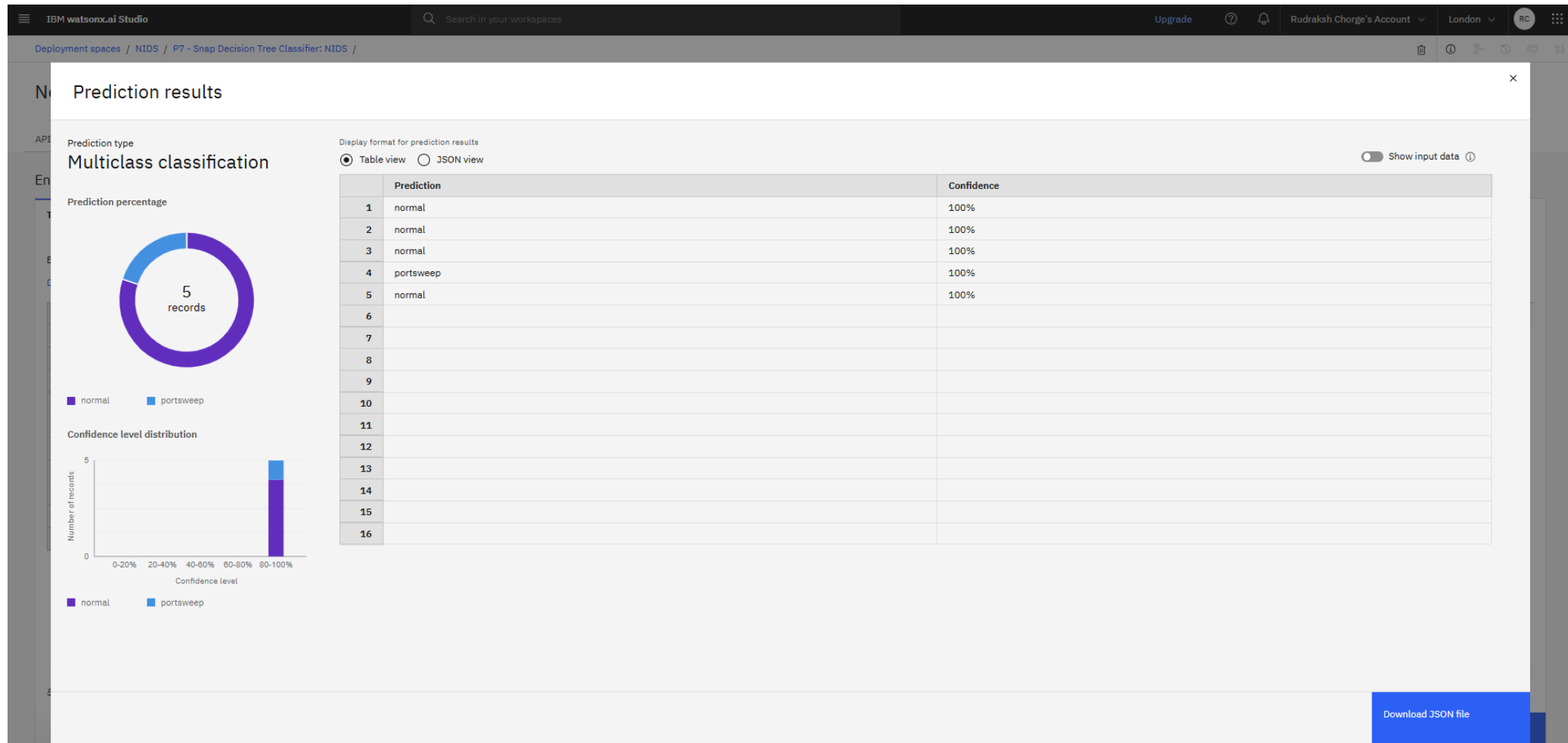
| | duration (double) | protocol_type (other) | service (other) | flag (other) | src_bytes (double) | dst_bytes (double) | land (double) | wrong_fragment (double) | urgent (double) | hot (double) | num_failed_logins (double) | logged_in (double) | num... |
|----|-------------------|-----------------------|-----------------|--------------|--------------------|--------------------|---------------|-------------------------|-----------------|--------------|----------------------------|--------------------|--------|
| 1 | 0 | tcp | ftp_data | SF | 491 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | tcp | http | SF | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 3 | 0 | tcp | telnet | SF | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 4 | 0 | tcp | private | S0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | tcp | ftp | SF | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| 6 | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | |

5 rows, 41 columns

Predict

Note: Only partial input shown here. Full input includes 42 features as per NSL-KDD dataset.

RESULT



CONCLUSION

- The ML model effectively distinguishes between normal and malicious network traffic.
- IBM AutoAI streamlined model creation and deployment.
- NSL-KDD dataset helped train a robust, multiclass model.
- The system can help automate detection in real-time monitoring tools.

FUTURE SCOPE

- Real-time intrusion detection with streaming data
- Integration with SIEM tools (e.g., Splunk, QRadar)
- Use of deep learning models (e.g., LSTM, CNN) for sequential analysis
- Handle encrypted packet features using advanced methods
- Deployment in **edge environments** (e.g., routers, IoT gateways) for low-latency detection
- Combining IDS with **threat intelligence platforms** for proactive defense
- Expanding the system to detect insider threats and anomalies in user behavior

REFERENCES

- NSL-KDD Dataset:
<https://www.kaggle.com/datasets/hassan06/nslkdd?select=KDDTrain%2B.txt>
- IBM Cloud and Watson Studio

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Rudraksh Chorge

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



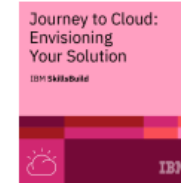
Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/5f6824b3-4a31-43ad-963e-2400a940bc3c>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Rudraksh Chorge

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/ed198219-649e-4460-a8ec-b1d8f68b4518>



IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to

Rudraksh Chorge

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