CAPSTONE PROJECT

NETWORK INTRUSION DETECTION SYSTEM USING MACHINE LEARNING ON IBM CLOUD

Presented By:
Shreya Langote
- Deogiri Institute of Engineering and Management Studies
- CSE(AIML)
- Github link: https://github.com/ShreyaLangote

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 vulnerable to various cyberattacks such as Denial-of-Service (DoS), Probe, Remote-to-Local (R2L), and User-to-Root (U2R). Detecting these attacks early is crucial to securing systems. This project aims to develop a machine learning-based Network Intrusion Detection System (NIDS) that can automatically identify and classify malicious activity from network traffic.

PROPOSED SOLUTION

- The proposed system aims to detect and classify anomalous behavior in network traffic using supervised machine learning. This is achieved using IBM Watson AutoAI to train and evaluate models, and Watson Machine Learning to deploy the best model as an online prediction service. The solution consists of the following components:
- **Data Collection:**
  - Use the Kaggle Network Intrusion Detection dataset
  - Data includes features like protocol type, service, bytes sent, flag, and attack label
- **Data Preprocessing:**
  - Encode categorical columns (e.g., protocol, service, flag)
  - Set target variable: class (normal or anomaly)
  - Split data for training and evaluation using AutoAI
- **Machine Learning Algorithm:**
  - Implement a supervised machine learning algorithm to classify network traffic as either normal or malicious (anomaly).
  - Watson AutoAI was used to automatically preprocess the data, select features, and train multiple classification models including Random Forest, XGBoost, and Logistic Regression.
  - The best-performing model was identified as Snap Random Forest Classifier, with high accuracy and optimized F1 score for the positive class

.

PROPOSED SOLUTION

- **Deployment:**
  - The selected model was saved, promoted to a deployment space, and deployed as an online REST API using IBM Watson Machine Learning.
  - This deployment enables real-time classification of network traffic. The model is accessible via a public endpoint, making it suitable for integration with monitoring tools or dashboards.
- **Evaluation:**
  - Assess the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or other relevant metrics.
  - The model selected for deployment demonstrated the best trade-off between detection quality and runtime efficiency, especially for detecting malicious traffic.
  - Result: The final deployed model successfully classifies network traffic into normal or anomaly categories

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the Network Intrusion Detection System (NIDS). This includes setting up cloud services, selecting appropriate tools, automating model training using AutoAI, and deploying the solution for real-time use.

- **System requirements**

  •IBM Cloud Lite Account
  •IBM Cloud Object Storage (for dataset upload)
  •IBM Watson Machine Learning (for model deployment)
  •Internet Browser (Chrome or Edge)
  •Kaggle Dataset: Network Intrusion Detection (CSV, 42 columns)
  •IBM Watson Studio (for AutoAI)

- **Library required to build the model**

  (Handled internally by AutoAI, but typically includes the following:)
  •pandas – data manipulation
  •scikit-learn – machine learning models
  •xgboost, lightgbm – gradient boosting classifiers
  •joblib – model serialization
  •IBM AutoAI Engine – AutoML pipeline construction
  •Watson Machine Learning API – deployment and API management

ALGORITHM & DEPLOYMENT

Algorithm Selection:

- The chosen algorithm for this binary classification task is the **Snap Random Forest Classifier**, selected automatically by IBM Watson AutoAI after evaluating multiple pipelines.
- Random Forest was chosen due to its high accuracy, robustness to overfitting, and ability to handle both numerical and categorical features common in network traffic data.

Data Input:

The input features include 41 columns from the Kaggle NIDS dataset, such as:

- duration, protocol_type, service, flag
- src_bytes, dst_bytes, wrong_fragment, etc.

Categorical columns were encoded internally by AutoAI, and the target label was the class column (normal vs. anomaly).

Training Process:

Watson AutoAI automatically:

- Encoded categorical variables
- Split the data into training and validation sets
- Performed cross-validation
- Tuned hyperparameters to optimize accuracy and F1-score

Each model pipeline was evaluated and ranked, and the best-performing pipeline was saved.

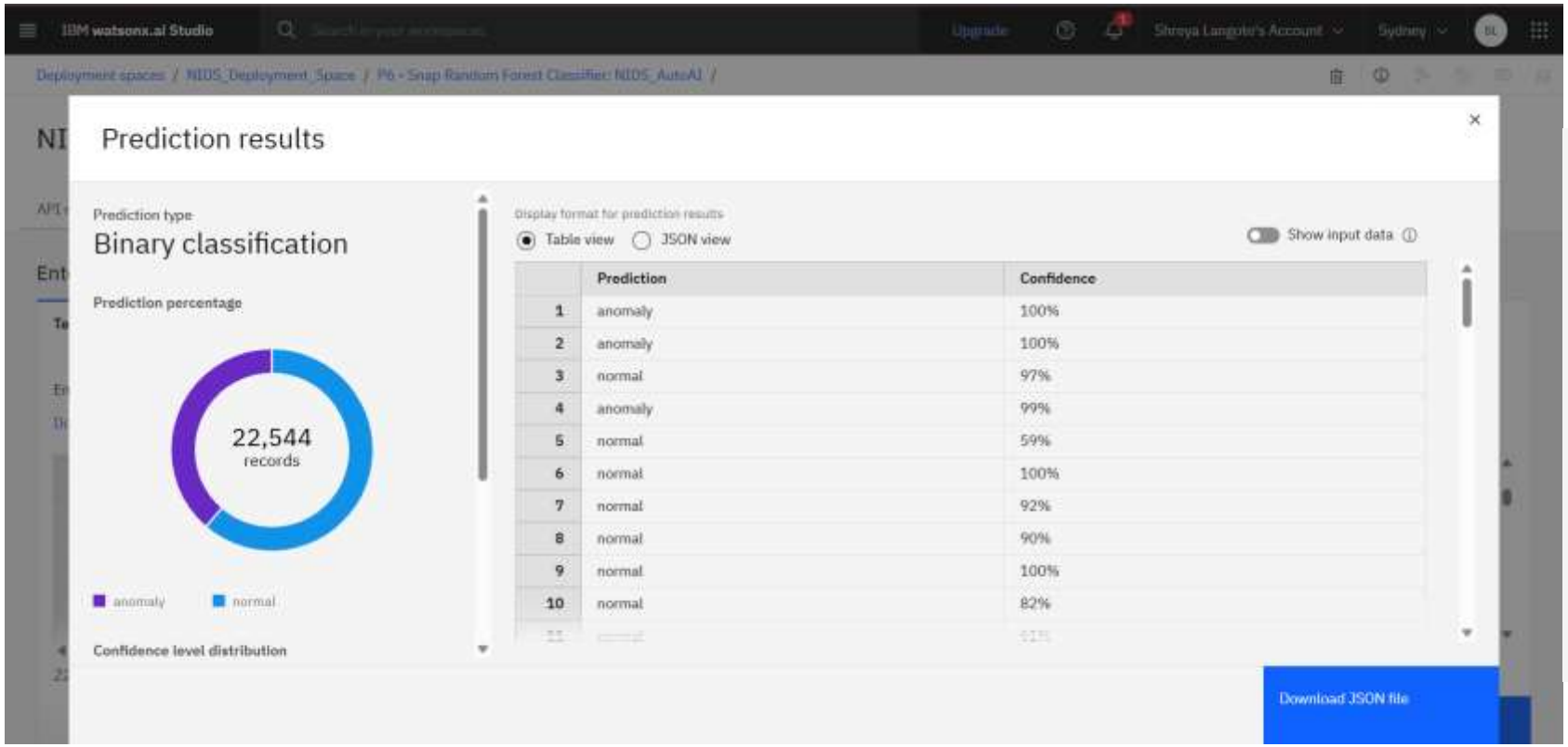
Prediction Process:

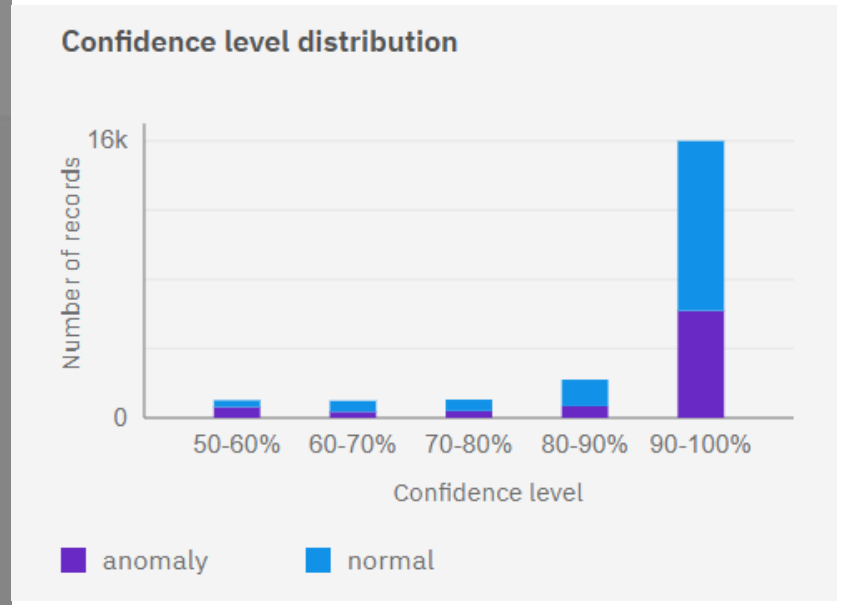
- The trained Random Forest model predicts whether a network activity instance is **normal** or **anomaly** based on real-time input data. Once deployed as a **REST API** via Watson Machine Learning, it accepts input in JSON format and returns predictions instantly. This setup enables the model to be integrated into real-world systems for live intrusion detection.

edunet foundation

RESULT

The machine learning model trained using IBM Watson AutoAI achieved strong performance in detecting anomalous network activity. The best pipeline, based on the Snap Random Forest Classifier, was selected for deployment after outperforming other algorithms in terms of accuracy and cross-validation metrics.





edunet foundation

CONCLUSION

- The proposed solution effectively demonstrates the use of machine learning to detect network intrusions by classifying network activity as either normal or anomalous. By leveraging IBM Watson AutoAI, the system automated data preprocessing, model selection, and training, resulting in a high-performing classification model. The model was successfully deployed as an online service using IBM Watson Machine Learning, enabling real-time predictions. The project validates the practicality of using cloud-based AutoML tools for building scalable and accurate intrusion detection systems with minimal manual intervention.

FUTURE SCOPE

Several enhancements and expansions can be made to improve the system's effectiveness and scalability:

- **Multiclass Classification**: Expand the binary classification into detailed intrusion types such as DoS, Probe, R2L, and U2R for more granular detection.
- **Real-Time Data Integration**: Connect the model with real-time network logs or packet capture tools to analyze live traffic.
- **Advanced Algorithms**: Incorporate deep learning techniques like LSTM or CNN to better capture sequential and spatial patterns in network behavior.
- **Edge Computing Deployment**: Deploy lightweight intrusion detection models on edge devices (routers, firewalls) to enable faster threat detection at the network boundary.
- **Cross-Domain Adaptability**: Extend the system to work across different organizational networks or cloud environments with domain adaptation techniques.
- **User Interface & Dashboard**: Build a user-friendly frontend to visualize alerts, logs, and prediction metrics in real time.

REFERENCES

1.Kaggle Dataset – Network Intrusion Detection
https://www.kaggle.com/datasets/sampadab17/network intrusion-detection

2.IBM Cloud Documentation – Watson Studio
https://www.ibm.com/cloud/watson-studio

3.IBM AutoAI User Guide
https://dataplatform.cloud.ibm.com/docs/content/wsj/autoai

4.IBM Watson Machine Learning Documentation
https://www.ibm.com/docs/en/wml

5.Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32.
https://doi.org/10.1023/A:1010933404324

6.Scikit-learn: Machine Learning in Python
https://scikit-learn.org/

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM

Shreya Langote

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/06925e99-db1c-4b1b-93b6-9b3b3e437f73

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Shreya Langote

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/a340337e-1168-4d4b-9312-86e074105d15

IBM

IBM CERTIFICATIONS

7/26/25, 7:16 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Shreya Langote

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

edunet foundation

THANK YOU

edunet foundation