CAPSTONE PROJECT

NETWORK INTRUSION DETECTION

Presented By:
1. Khushboo Pandey-Shri Ramswaroop Memorial University-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

- With the rise in cyber-attacks targeting modern digital infrastructures, there is a growing need for intelligent systems capable of identifying and mitigating security threats in real time. The objective of this project is to develop a Network Intrusion Detection System (NIDS) using IBM Cloud Pak for Data that can accurately distinguish between normal and anomalous network traffic. Using a labeled dataset sourced from Kaggle, which includes both training and testing data, the system leverages machine learning to analyze traffic patterns and detect various types of attacks such as DoS, Probe, R2L, and U2R. The end goal is to enhance the security of communication networks by providing early warnings of potential intrusions.

PROPOSED SOLUTION

- **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.**
- **The solution consists of the following components:**
- **Data Collection:** The dataset, downloaded from Kaggle, includes labeled network traffic instances categorized as normal or one of several attack types.
- **Data Preprocessing:** Preprocessing steps involved data cleaning, handling missing values, encoding categorical features, and normalization to prepare the data for model training.
- **Machine Learning Algorithm:** Multiple ML algorithms were tested and compared; the best-performing model Snap Decision Tree Classifier was selected based on accuracy and cross validation score. The model's performance was validated using the separate test set, ensuring its effectiveness in real-world scenarios.
- **Deployment:** The trained model was deployed on IBM Cloud Pak for Data as a real-time inference service to monitor live network traffic.
- **Evaluation:** A monitoring dashboard was implemented to visualize predictions and generate alerts upon detection of suspicious or anomalous activity.

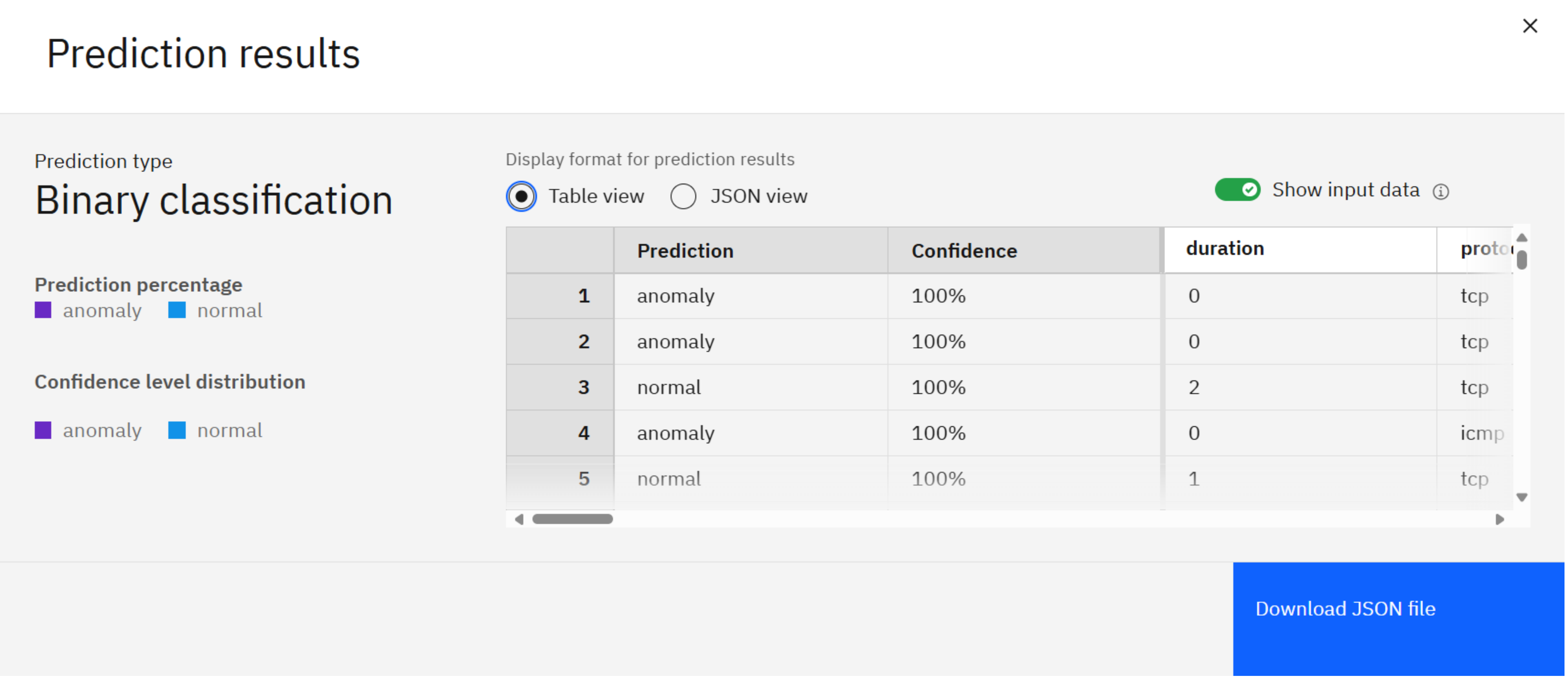
SYSTEM APPROACH

- **System requirements:**
  - **IBM Cloud Pak for Data** (cloud environment for building and deploying Machine Learning models)
  - Minimum **8 GB RAM** (16 GB recommended for large datasets)
  - Stable internet connection for cloud integration and dataset access

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  - The **Snap Decision Tree Classifier** was selected for its high accuracy and speed, enhanced by **HPO-1** (Hyperparameter Optimization Level 1). It efficiently handles large, complex network traffic data and offers better performance than standard classifiers for intrusion detection.
- **Data Input:**
  - The model uses features like protocol type, service, source/destination bytes, connection counts, and statistical attributes to classify traffic as normal or one of several attack types (DoS, Probe, R2L, U2R).
- **Training Process:**
  - The model was trained on a Kaggle dataset after preprocessing steps such as encoding, normalization, and cleaning. HPO-1 tuning and stratified cross-validation improved accuracy and balanced class representation.
- **Prediction Process:**
  - Deployed on **IBM Cloud Pak for Data**, the trained model processes live or batch traffic data and instantly classifies it as normal or anomalous, supporting real-time intrusion alerts.

RESULT



Prediction results

Prediction type

Binary classification

Prediction percentage

🟪 anomaly 🟦 normal

Confidence level distribution

🟪 anomaly 🟦 normal

Display format for prediction results

● Table view ○ JSON view

Show input data

	Prediction	Confidence	duration	proto
1	anomaly	100%	0	tcp
2	anomaly	100%	0	tcp
3	normal	100%	2	tcp
4	anomaly	100%	0	icmp
5	normal	100%	1	tcp

CONCLUSION

- **Findings & Discussion**
  - The proposed Network Intrusion Detection System using the **Snap Decision Tree Classifier** on **IBM Cloud Pak for Data** successfully identified normal and anomalous network traffic with high accuracy. The model effectively detected various attack types (DoS, Probe, R2L, U2R), demonstrating strong generalization across imbalanced classes due to optimized preprocessing and HPO-1 tuning.
  - The system's real-time prediction capabilities make it a practical tool for early warning and prevention of cyber threats, enhancing overall network security.
- **Challenges Faced**
  - **Data imbalance**, especially with rare attack types like U2R and R2L, affected initial model performance.
  - **Feature selection** and preprocessing required extensive tuning to reduce noise and improve model precision.
- **Potential Improvements**
  - Implementing **ensemble models** or **deep learning techniques** for further accuracy gains.
  - Incorporating **real-time traffic streams** for dynamic learning and detection.
  - Using **auto-scaling deployment** to handle high network loads efficiently.
- **Conclusion**
  - Accurate and timely detection of network intrusions is critical for safeguarding digital infrastructure. This project demonstrates that an ML-based approach, when properly optimized and deployed, can significantly improve intrusion detection and response, reducing the risk of cyber-attacks and data breaches.

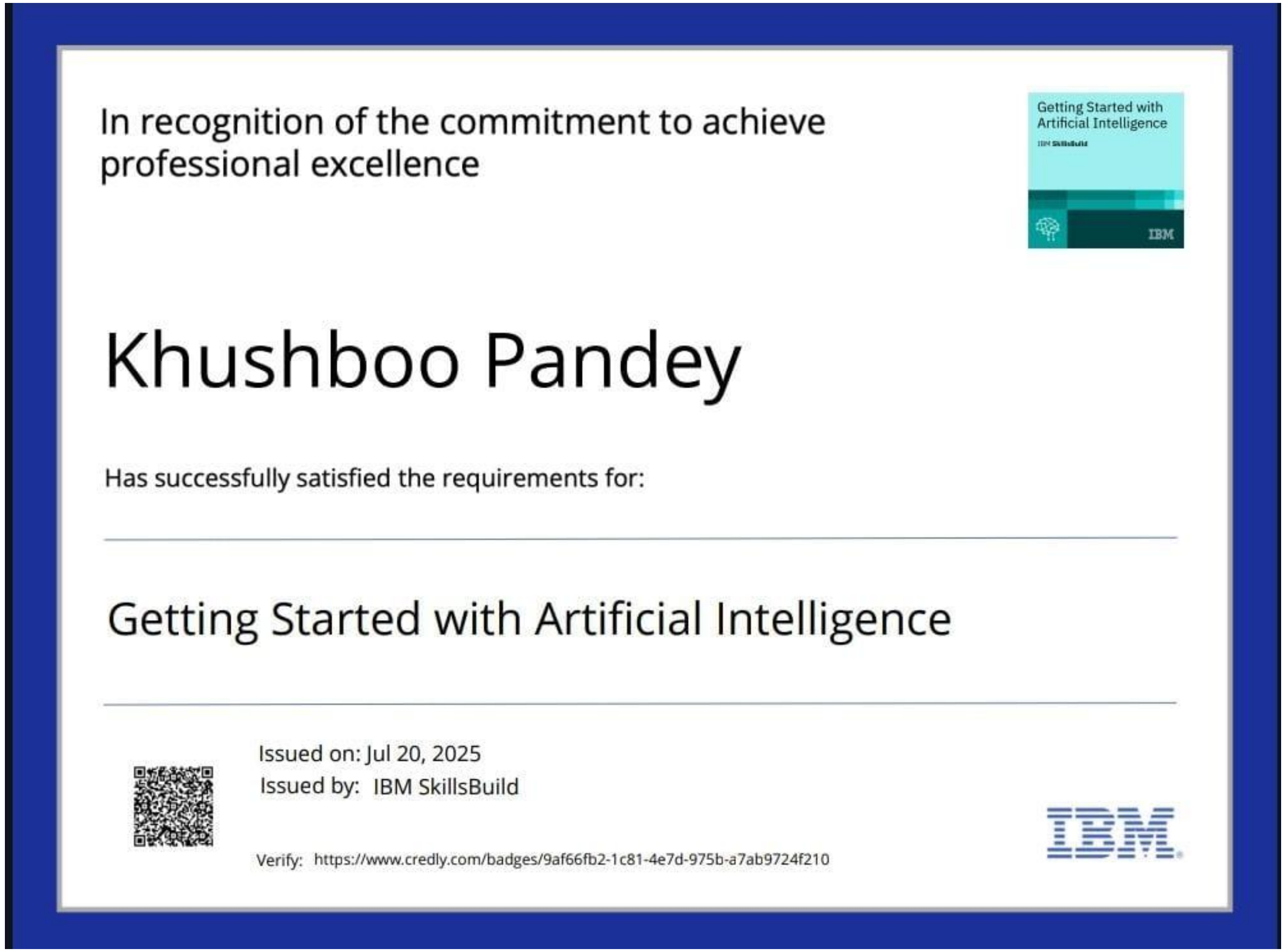
FUTURE SCOPE

- To further enhance and expand the Network Intrusion Detection System, several improvements can be considered. Incorporating additional data sources such as real-time traffic from firewalls, routers, and external threat intelligence feeds can provide deeper insights and improve the detection of diverse attack vectors.
- Optimizing the algorithm through advanced hyper-parameter tuning techniques or adopting ensemble methods like Random Forest combined with Gradient Boosting can lead to more accurate and robust classifications.
- Expanding the system’s coverage across multiple cities or organizational branches would allow for centralized monitoring while tailoring detection parameters to regional traffic behaviors.
- Moreover, integrating emerging technologies such as edge computing can enable faster, localized detection at the network edge, reducing response time. Advanced techniques like deep learning (e.g., LSTM for temporal patterns) or federated learning can also be leveraged for better adaptability and privacy-preserving training. These enhancements will not only improve performance but also future-proof the system against evolving cybersecurity threats.

REFERENCES

- Key Dataset Reference
  - **Sampada Bhosale.** *Network Intrusion Detection Dataset*, Kaggle, ~768 KB, contains a variety of simulated intrusions in a military network environment, labeled for anomaly detection (e.g., DoS, Probe, R2L, U2R) kaggle.com+9kaggle.com+9kaggle.com+9.
- Supporting Literature & Research Papers
  - **Chandola, V., Banerjee, A., & Kumar, V.** *Anomaly Detection: A Survey*. ACM Computing Surveys, 2009. A foundational survey on anomaly detection techniques in networks and cybersecurity contexts.
  - **Moustafa, N. & Slay, J.** *UNSW-NB15: A Comprehensive Dataset for Network Intrusion Detection Systems*. IEEE MilCIS, 2015. Introduces a modern benchmark dataset used in many intrusion detection studies.
  - **Psychogyios, K. et al.** *Deep Learning for Intrusion Detection Systems (IDSs) in Time Series Data*. Future Internet, 2024. Demonstrates proactive IDS modeling using UNSW-NB15 in a time-series format. mdpi.com+1researchgate.net+1
  - **Alsariera, Y. A.** *Detecting Generic Network Intrusion Attacks Using Tree-based Machine Learning Methods*. Explores decision-tree algorithms (J48, Random Tree) applied to intrusion detection classification tasks. researchgate.net
  - **He, K., Kim, M.-R., & Asghar, M. R.** *Adversarial Machine Learning for Network Intrusion Detection Systems: A Comprehensive Survey*. IEEE Communications Surveys & Tutorials, 2023. Reviews challenges in deploying ML-based IDS under adversarial conditions.

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

Khushboo Pandey

Has successfully satisfied the requirements for:

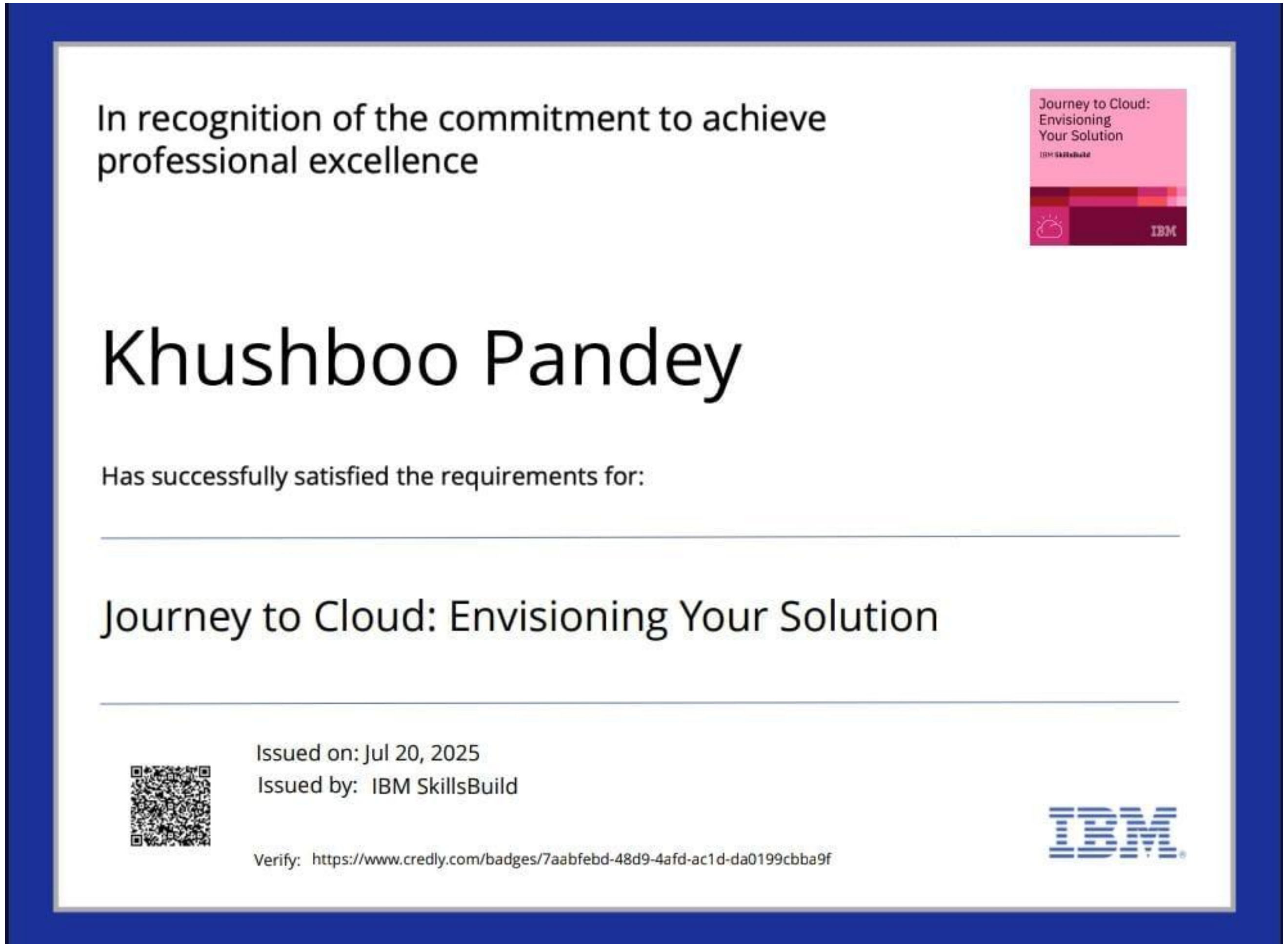
Getting Started with Artificial Intelligence

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/9af66fb2-1c81-4e7d-975b-a7ab9724f210

IBM

IBM CERTIFICATIONS



Journey to Cloud:
Envisioning
Your Solution
IBM SkillsBuild
IBM

In recognition of the commitment to achieve professional excellence

Khushboo Pandey

Has successfully satisfied the requirements for:

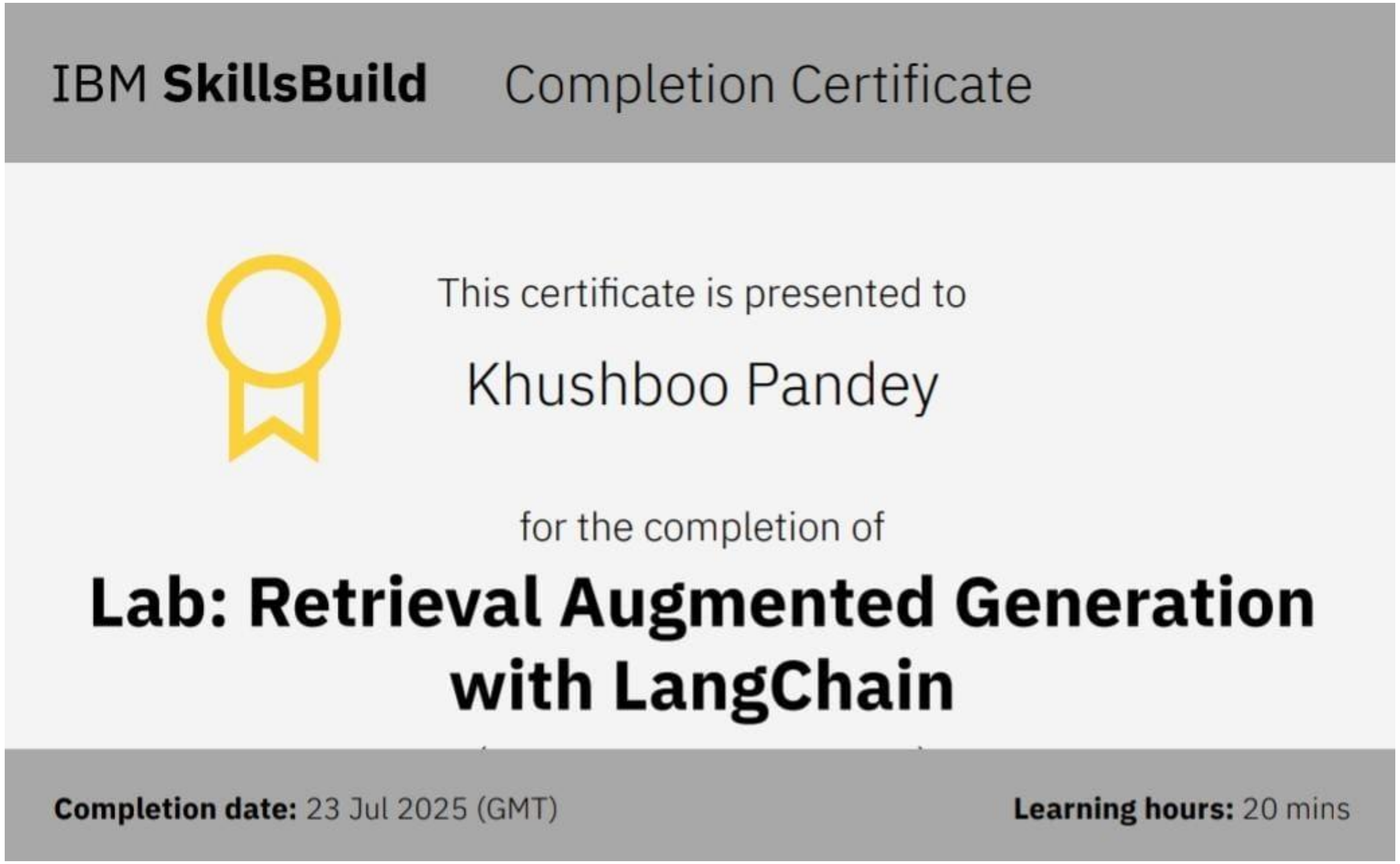
Journey to Cloud: Envisioning Your Solution

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/7aabfebd-48d9-4afd-ac1d-da0199cbba9f

IBM

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Khushboo Pandey

for the completion of

Lab: Retrieval Augmented Generation with LangChain

Completion date: 23 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU