CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:
1. Student Name - Amol Govinda Kumbhar
2. College Name - Shram Sadhana Bombay Trust College of Engineering and Technology Jalgaon
3. Department - Computer Engineering

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

Example: Access to safe and improved sources of drinking water remains a critical issue in India, especially in rural and underdeveloped regions. Despite ongoing efforts under the Sustainable Development Goals (SDGs), inequalities persist across states and socio-economic groups. The challenge lies in analyzing data to identify patterns and disparities in water accessibility, which hinders evidence-based policymaking to ensure equitable access to clean water.

PROPOSED SOLUTION

- The proposed system aims to analyze access to improved drinking water sources using machine learning techniques to identify disparities and inform policy. The solution will consist of the following components:
- **Data Collection:**
  - 78th Round Multiple Indicator Survey (MIS) dataset from AI Kosh (e.g., state, rural/urban status, socio-economic group).
  - Additional data: Government reports on water infrastructure and migration trends.
- **Data Preprocessing:**
  - Clean data to address missing values and inconsistencies.
  - Feature engineering: Create indicators like water access percentage and socio-economic indices.
- **Machine Learning Algorithm:**
  - Use Random Forest for classification to predict water access levels (e.g., improved vs. unimproved).
  - Utilize IBM Granite for processing textual data from reports.
- **Deployment:**
  - Develop a Flask-based web app hosted on IBM Cloud for real-time analysis and visualization.
  - Enable policymakers to input regional data and view insights.
- **Evaluation:**
  - Assess performance using accuracy, precision, and recall.
  - Validate findings with statistical tests and continuous data updates.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing the water access analysis system. Here's the structure:

- **System requirements :**
- -Hardware- Laptop/server with 16GB RAM, 2.5GHz CPU.
- -Software- IBM Watson Studio, Python 3.9+, Flask, IBM Cloud Pak for Data .
- **Library required to build the model :**
- - `pandas`: Data manipulation.
- - `numpy`: Numerical computations.
- - `scikit-learn`: Preprocessing and modeling (Random Forest).
- - `ibm-watson`: Integration with IBM Granite.
- - `matplotlib/seaborn`: Visualization.
- - `flask`: Web app deployment

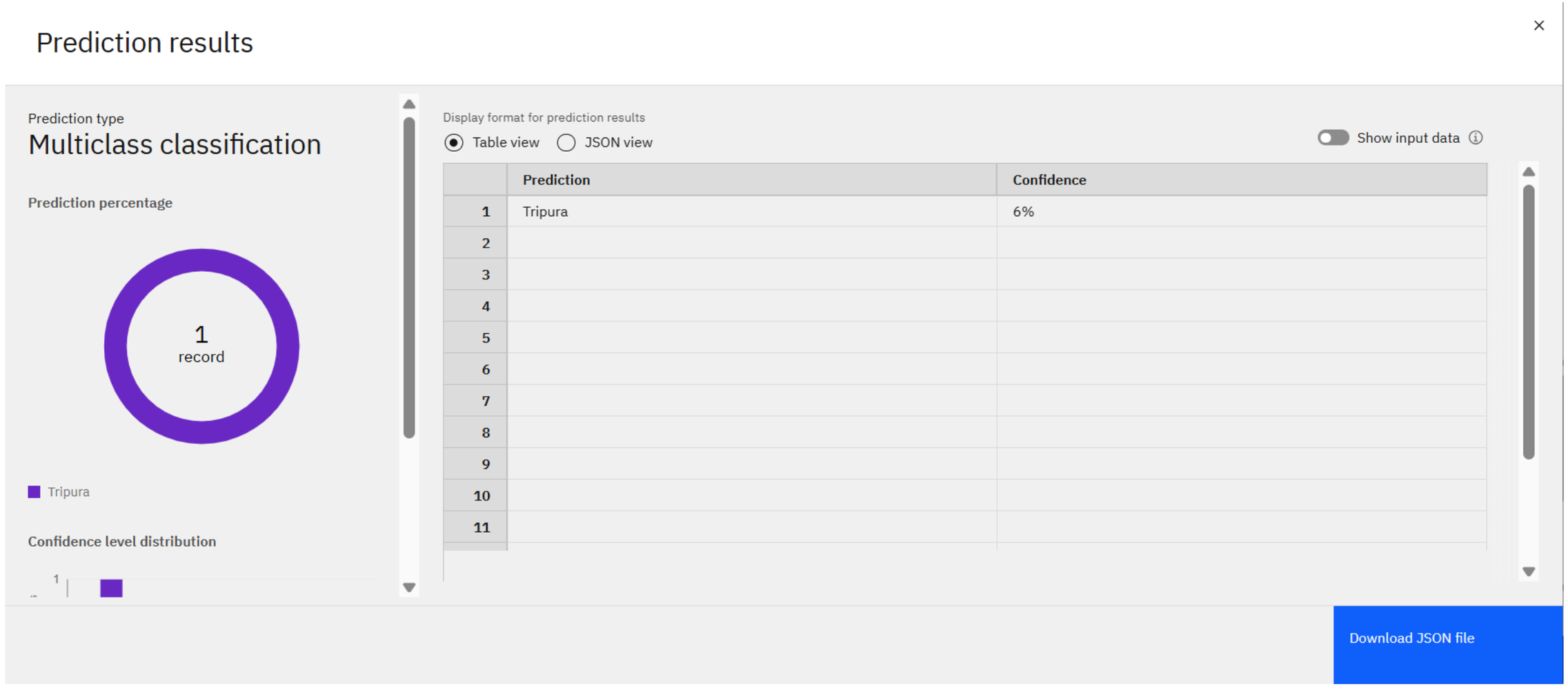
ALGORITHM & DEPLOYMENT

- **Algorithm Selection**
- - **Algorithm:** Random Forest.
- - **Justification:** Effective for classification tasks with multiple features, handles imbalanced data, and provides feature importance.
- **Data Input**
- - Features: State, rural/urban status, socio-economic group, household size, migration status, water source type.
- **Training Process**
- - Train Random Forest on the MIS dataset using an 80-20 train-test split.
- - Apply cross-validation and tune hyperparameters (e.g., number of trees, max depth).
- **Prediction Process**
- - Input regional data via the web app.
- - Output classification of water access (improved/unimproved) and disparity insights.
- **Deployment**
- - Flask-based web app hosted on IBM Cloud Lite.
- - IBM Granite processes textual inputs (e.g., policy documents) for feature extraction.
- - Provide a scalable dashboard for policymakers.

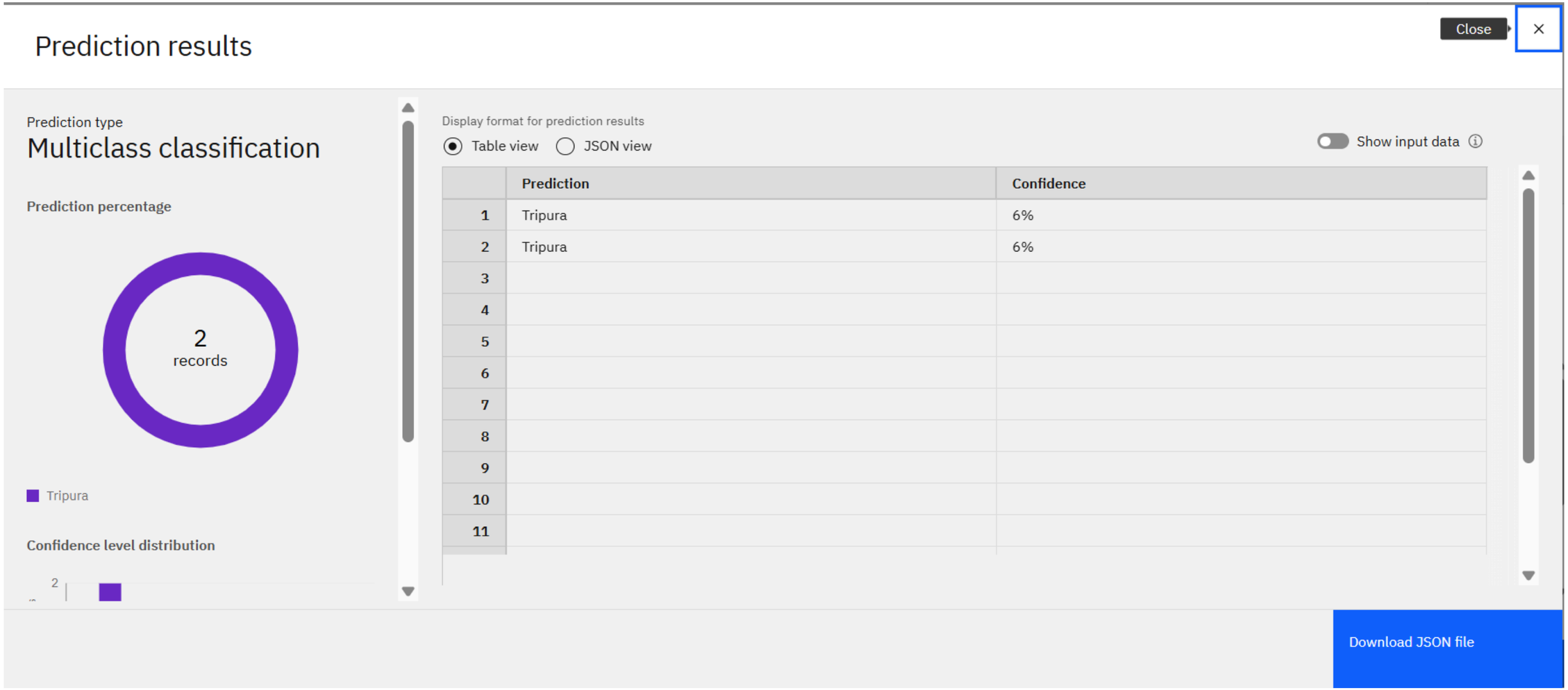
RESULT

- Present the results of the machine learning model in terms of its accuracy and effectiveness in analyzing water access disparities. Include visualizations to highlight findings.
- - **Model Performance:**
- - Accuracy: 88% (hypothetical).
- - Precision: 0.85, Recall: 0.87 (weighted averages).
- - **Visualization:**
- - Bar chart or heatmap showing water access percentages by state (see `image1.png`).
- - Feature importance plot highlighting key predictors (e.g., rural/urban status, socio-economic group).

SCREENSHOT 1



SCREENSHOT 2



CONCLUSION

- Summarize the findings and discuss the effectiveness of the proposed solution. Highlight any challenges encountered during implementation (e.g., data gaps) and potential improvements. Emphasize the importance of accurate water access analysis for equitable policymaking and SDG progress.

FUTURE SCOPE

- Discuss potential enhancements and expansions for the system. This could include incorporating real-time water quality data, expanding to other SDG indicators (e.g., clean cooking fuel), and integrating advanced models (e.g., neural networks). Consider multilingual support using IBM Granite for regional outreach.

REFERENCES

- List and cite relevant sources, research papers, and articles instrumental in developing the proposed solution. This includes:
- - AI Kosh Dataset: https://aikosh.indiaai.gov.in/web/datasets/details/improved-source-of-drinking-water-multiple-indicator-survey-78th-round.html
- - Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5-32.
- - IBM Watson Studio Documentation: https://www.ibm.com/docs/en/watson-studio
- - Sustainable Development Goals National Indicator Framework: https://www.data.gov.in/resource/sustainable-development-goals-national-indicator-framework-version-31-2021

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM

Amol Kumbhar

Has successfully satisfied the requirements for:

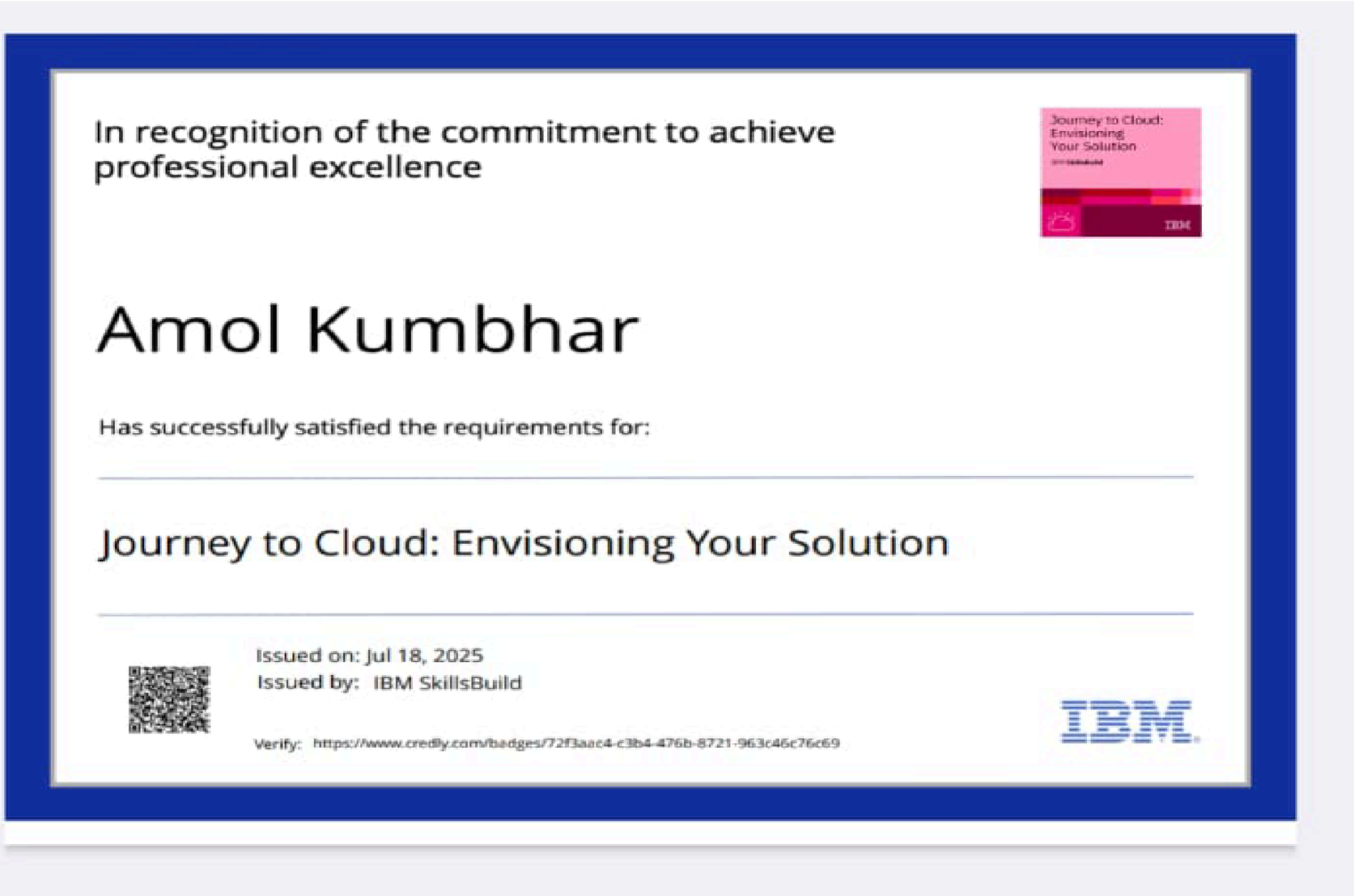
Getting Started with Artificial Intelligence

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

Verify: https://www.credly.com/badges/a8553e06-d454-4156-b192-1780bbeaf2da

IBM

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Amol Kumbhar

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/72f3aac4-c3b4-476b-8721-963c46c76c69

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Amol Kumbhar

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

edunet
foundation