CAPSTONE PROJECT

DRINKING WATER ACCESS PREDICTION

Presented By:
Student Name: Phani Bhaskara Rama Krishna Reddy Padala
College Name: Lovely Professional University
Department: Computer Science and Engineering

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

- Access to safe drinking water is a basic human necessity, yet millions of people worldwide still rely on unimproved sources, leading to waterborne diseases and poor health outcomes.
- There is a need for a reliable prediction system that can classify whether a water source is improved or unimproved based on various environmental, demographic, and infrastructure-related factors.
- Accurate prediction can help policymakers and communities prioritize interventions for clean water access, ultimately reducing inequalities and supporting sustainable development goals.

PROPOSED SOLUTION

- Use AI Kosh dataset (78th MIS) containing state-wise demographic, migration, and utility indicators.
- Develop a regression model to predict drinking water access.
- Deploy model using IBM Watson Studio for real-time prediction.
- Provide interactive UI to test predictions with custom inputs.
- Support policymakers with data-driven decisions.

SYSTEM APPROACH

This section outlines the overall strategy, tools, and methodology used to develop and deploy a machine learning model that predicts the percentage of the population with access to improved drinking water sources across different regions using AutoAI.

System Requirements:

Component	Specification
Device	Laptop/Desktop with minimum 4GB RAM (8GB+ recommended for smoother operation)
Internet Speed	Stable internet connection (minimum 10 Mbps recommended)
Browser	Google Chrome / Mozilla Firefox (latest version recommended)
Cloud Environment	IBM Cloud Academic Portal
Account Access	IBM Cloud account with access to Watsonx.ai and Cloud Object Storage
Platform	IBM Watsonx.ai (AutoAI)
Runtime	Watsonx.ai Runtime service (provisioned for model building and deployment)
Deployment	AutoAI-generated model deployed in a Watsonx.ai Deployment Space
Dataset Format	CSV (Improved source of Drinking Water.csv)

SYSTEM APPROACH

Although IBM AutoAI handles model creation without manual coding, the backend utilizes several essential libraries and frameworks for data processing, model training, and evaluation:

Libraries Required:

Library/Tool	Purpose
pandas	Data manipulation and analysis
numpy	Numerical computations
scikit-learn	Machine learning models, preprocessing, pipeline creation
xgboost / lightgbm	Advanced tree-based boosting algorithms (AutoAI uses automatically)
watson-machine-learning-client	Interacting with deployed models on IBM Cloud

Note: All of these are handled automatically within the AutoAI pipeline; no manual installation or coding is required unless additional customization is needed outside the AutoAI interface.

ALGORITHM & DEPLOYMENT

ALGORITHM OVERVIEW

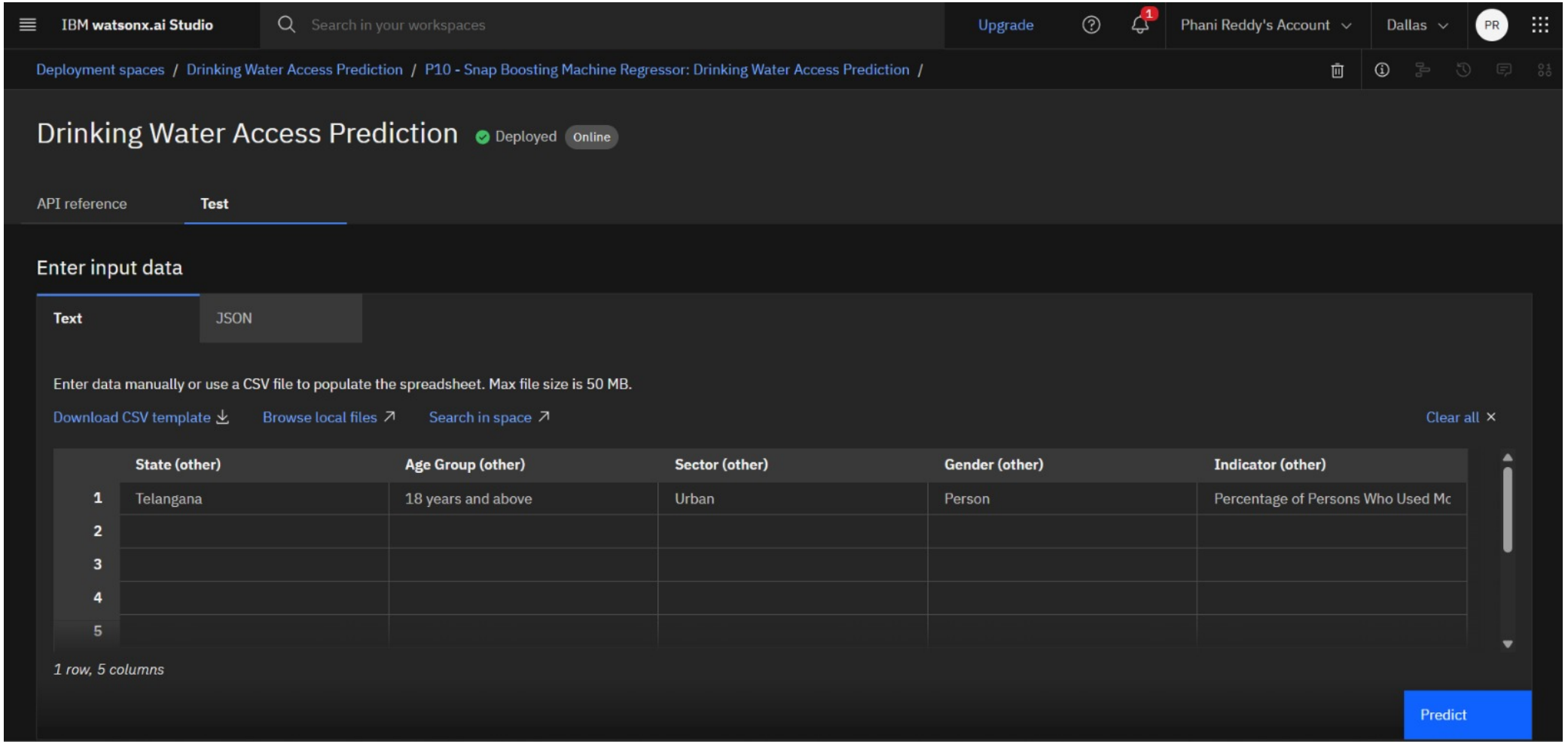
- This project uses **IBM Watsonx.ai AutoAI**, which automates model training and selection. It performs:
- **Automatic preprocessing**: Handles missing values, data types, and feature engineering.
- **Model generation**: Builds multiple pipelines using algorithms like:
  - Random Forest Regressor
  - Gradient Boosting (XGBoost/LightGBM)
  - Decision Tree, Linear, and Ridge/Lasso Regression
- **Evaluation**: Ranks models using metrics like **R^2**, **MAE**, and **RMSE**.
- **Selection**: Chooses the best-performing pipeline for deployment.

ALGORITHM & DEPLOYMENT

Model Deployment

- After identifying the best-performing pipeline, the model is **deployed to the cloud** using IBM Watsonx.ai's integrated tools:
- **Model Promotion**:
  - The selected model is **promoted to a deployment space** within Watsonx.ai.
- **Deployment Creation**:
  - A **new deployment** is created (real-time or batch) to make the model accessible via API.
  - The deployment is named and version-controlled.
- **Model Testing**:
  - The deployed model is tested using **new input values** directly from the cloud UI.
  - The predicted output (i.e., the percentage of the population with access to improved drinking water sources) is generated in real time.

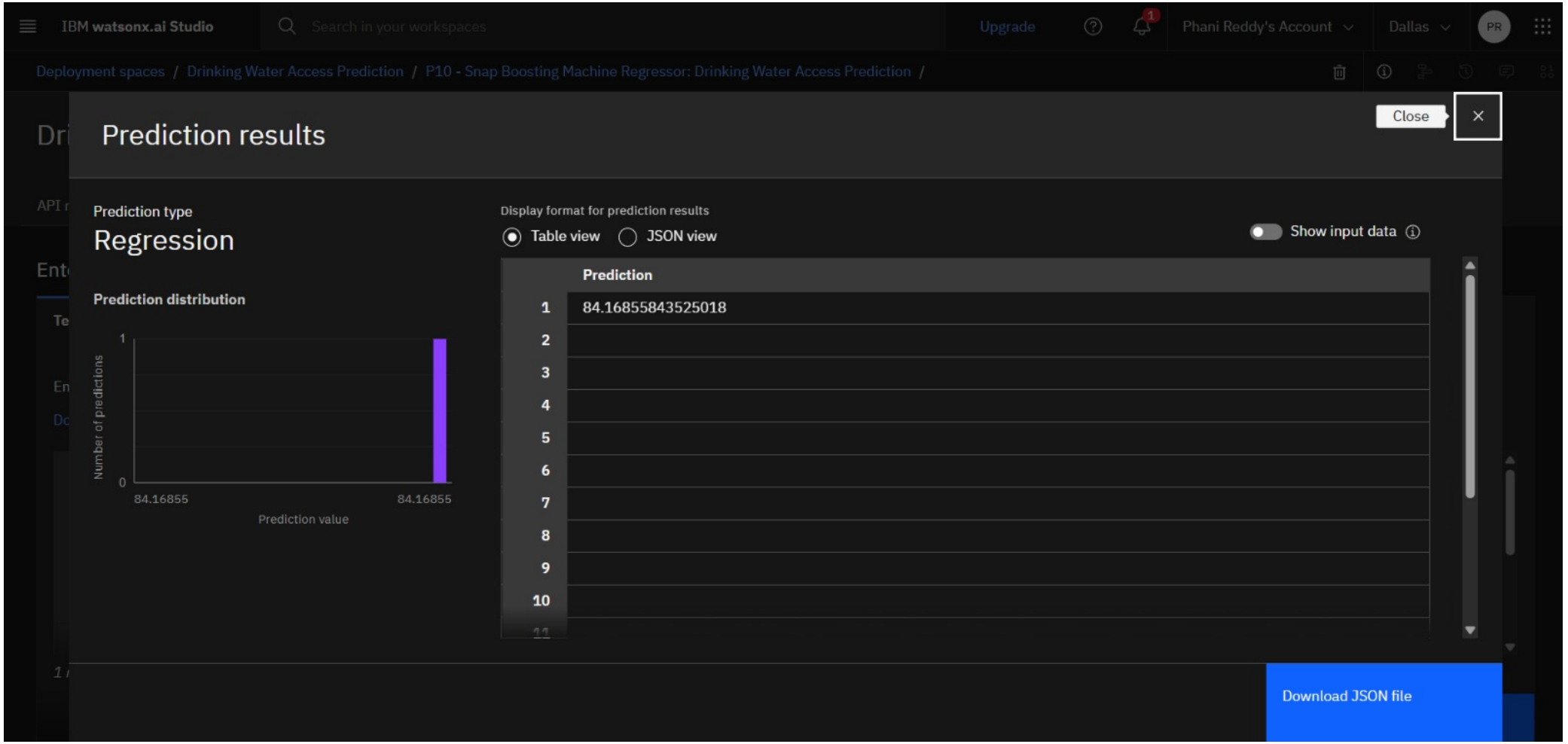
RESULT



INPUT DATA

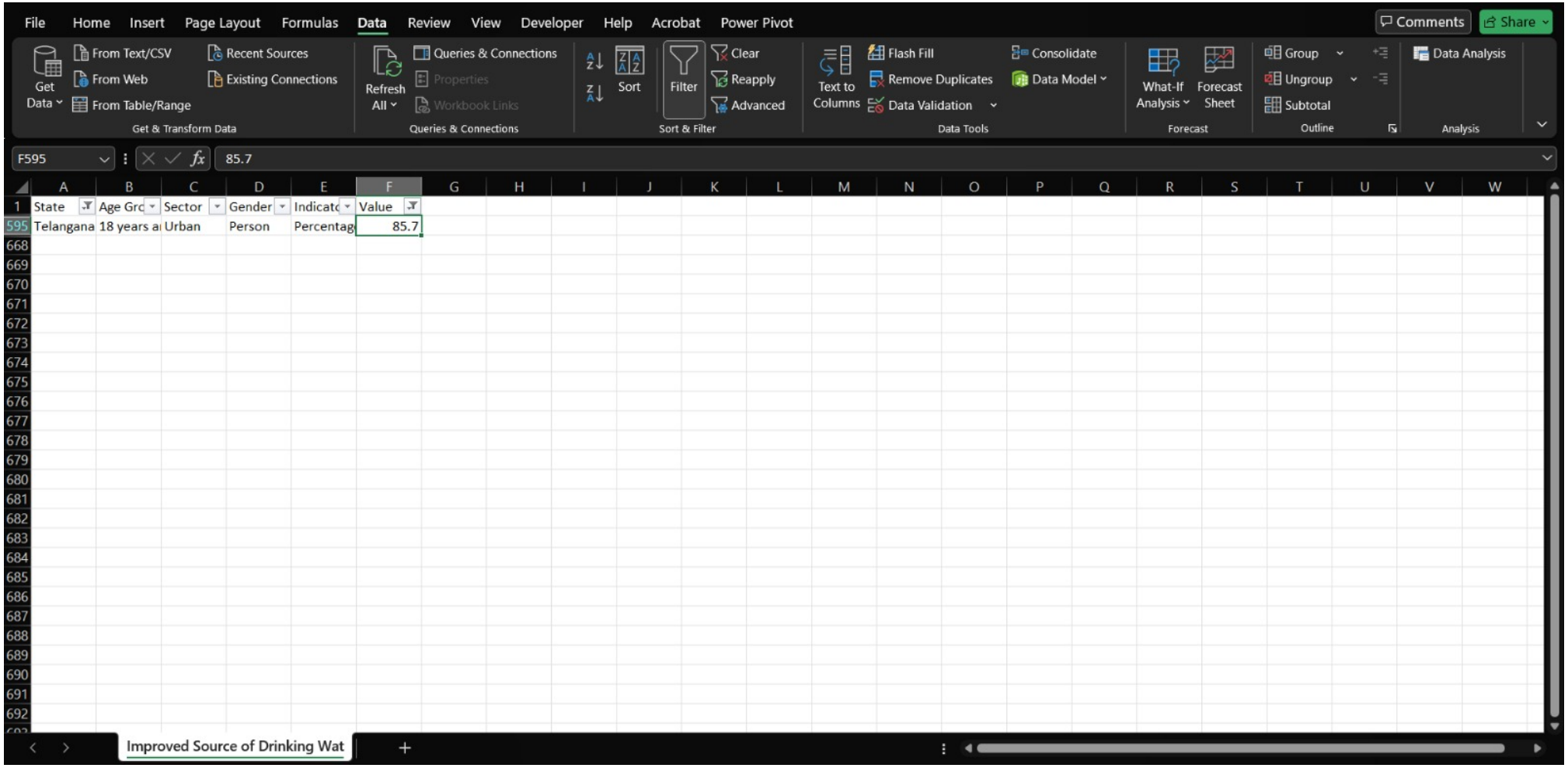
edunet foundation

RESULT



PREDICTION RESULT

RESULT



MODEL PREDICTION ACCURACY: COMPARISON WITH ACTUAL DATA

edunet foundation

CONCLUSION

- This project effectively used AutoAI to develop a predictive model for estimating **the percentage of the population with access to improved drinking water sources** across different regions of India. By leveraging inputs like state/UT, sector, age group, gender, and household indicators, the model uncovers key accessibility patterns and highlights regions with possible disparities in safe water access.
- AutoAI streamlined the entire process, from **data preprocessing** to **model selection and evaluation**, enabling the creation of a high-performing regression model with minimal manual effort. Additionally, future efforts could focus on deploying the model in user-friendly decision-support tools, enabling policymakers and local authorities to make data-driven interventions more efficiently and effectively..
- Moving forward, this predictive system can be enhanced by integrating **real-time data updates**, incorporating **socio-economic indicators**, and expanding the feature set to include factors like literacy rate or digital accessibility.

FUTURE SCOPE

- IoT Integration – Connect real-time water quality sensors for live monitoring and improved prediction accuracy.
- Climate & Seasonal Trends – Incorporate rainfall, temperature, and seasonal changes to refine accessibility forecasts.
- Geospatial Analysis – Use satellite imagery and GIS mapping to identify high-risk water scarcity regions.
- Policy Recommendation Engine – Generate region-specific strategies for better water resource management.
- Community Feedback Loop – Collect local user feedback to validate predictions and improve the model over time.
- Scalability Across Regions – Extend the model to other states or countries facing similar water access challenges.

REFERENCES

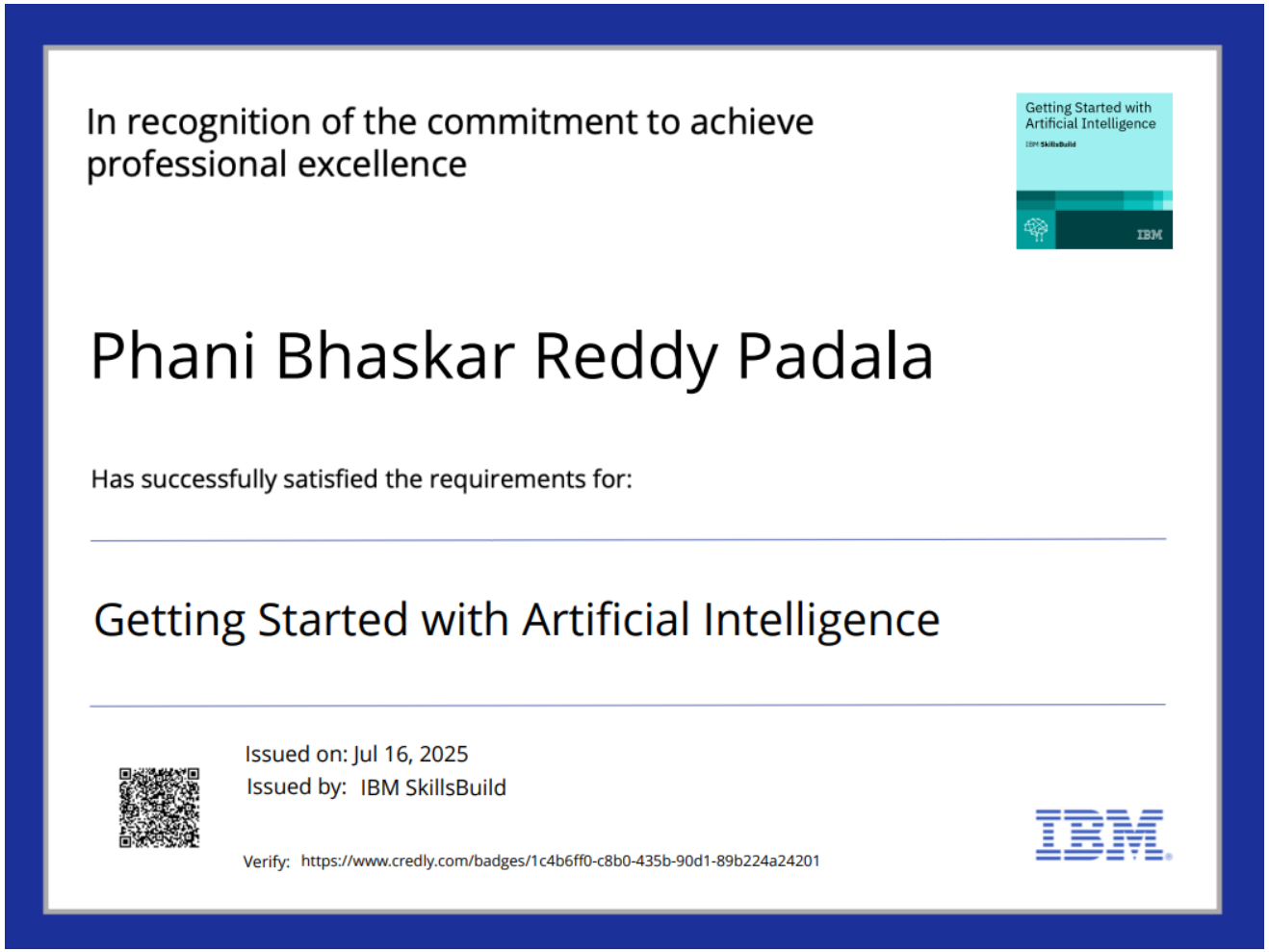
- Official Dataset Source: https://aikosh.indiaai.gov.in/web/datasets/details/improved_source_of_drinking_water_multiple_indicator_survey_78th_round.html
- IBM Cloud Platform: https://cloud.ibm.com/
- IBM Cloud: *Watsonx.ai Studio (AutoAI Tool Documentation)*

GITHUB LINK

https://github.com/PhaniBhaskarReddyPadala/IBM-EdunetFoundation-Project

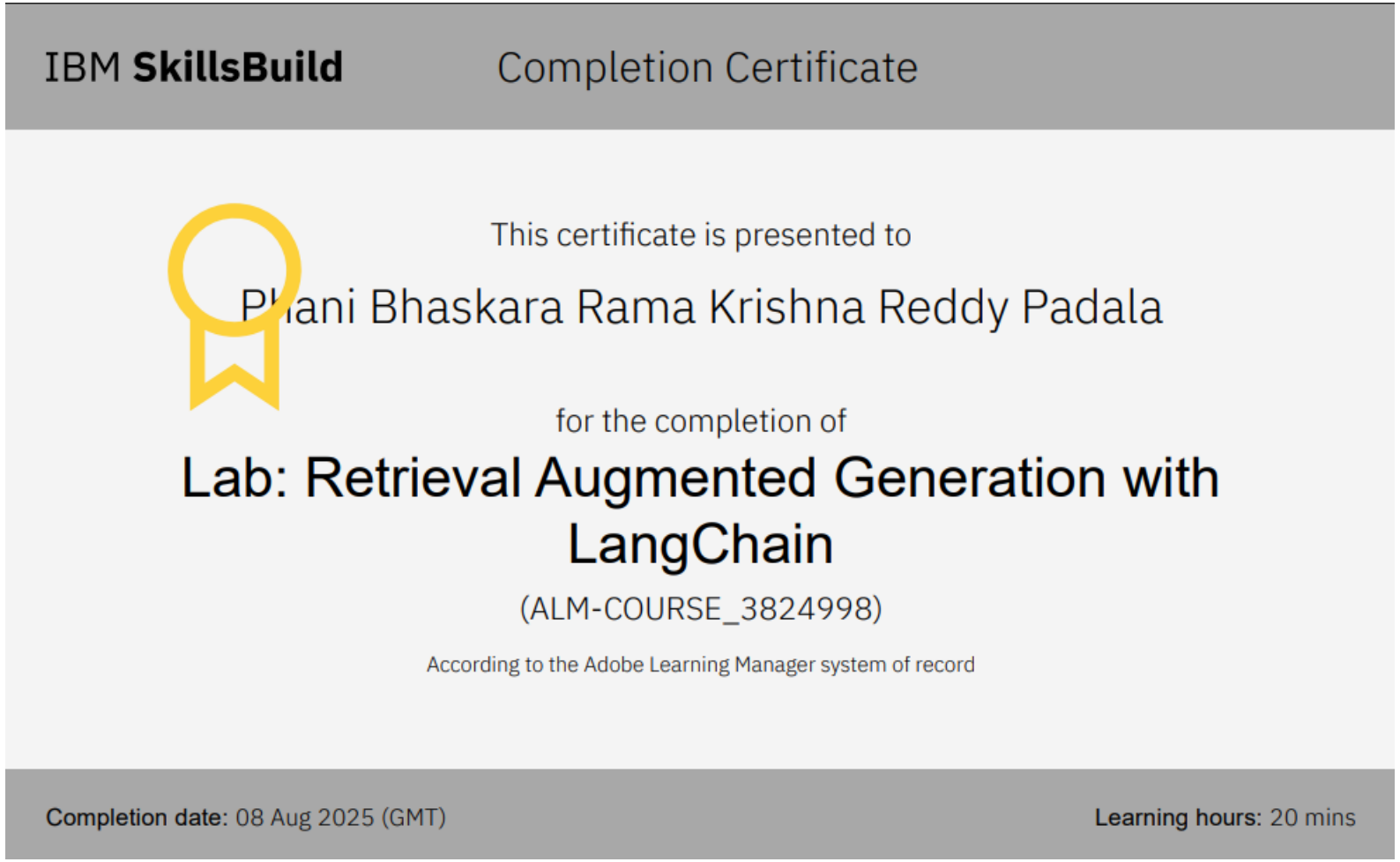
edunet foundation

IBM CERTIFICATIONS



IBM CERTIFICATIONS

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Phani Bhaskara Rama Krishna Reddy Padala

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 08 Aug 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet foundation