CAPSTONE PROJECT

ESTIMATING PROJECT COMPLETION METRICS FOR PMGSY SCHEMES

Presented By:
Student Name: Jaipal Reddy Karri
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

- Rural connectivity plays a vital role in economic growth and social development in India.
  The Pradhan Mantri Gram Sadak Yojana (PMGSY) aims to provide all-weather road connectivity to unconnected habitations.
  With multiple schemes like PMGSY-I, PMGSY-II, and RCPLWEA, classification of projects is essential for monitoring and planning.
- Manual classification of thousands of rural infrastructure projects is slow and error-prone.
  Each scheme has different objectives, funding structures, and technical specifications.
  Automating this process can improve transparency, efficiency, and decision-making.
- Machine learning can use project characteristics like length, cost, and location for classification.
  This enables quick identification of the scheme for each project.
  Accurate classification supports better budget allocation and policy evaluation.

PROPOSED SOLUTION

- Use the Al Kosh PMGSY dataset containing physical and financial characteristics of road and bridge projects.
- Clean and preprocess the dataset, handling missing values and standardizing feature formats.
- Develop a regression model in IBM Watson Studio (AutoAI) to analyze key numerical project metrics and assist in classification.
- Combine regression outputs with classification algorithms to accurately assign each project to its correct PMGSY scheme.
- Deploy the trained model on IBM Cloud Lite for real-time classification and prediction.
- Provide an interactive dashboard for users to test predictions, visualize results, and gain scheme-level insights.

SYSTEM APPROACH

- This section outlines the overall strategy, tools, and methodology used to develop and deploy a machine learning model that analyzes maternal health indicators and predicts factors influencing maternal mortality rates 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 (Estimating Project Completion Metrics for PMGSY Schemes.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

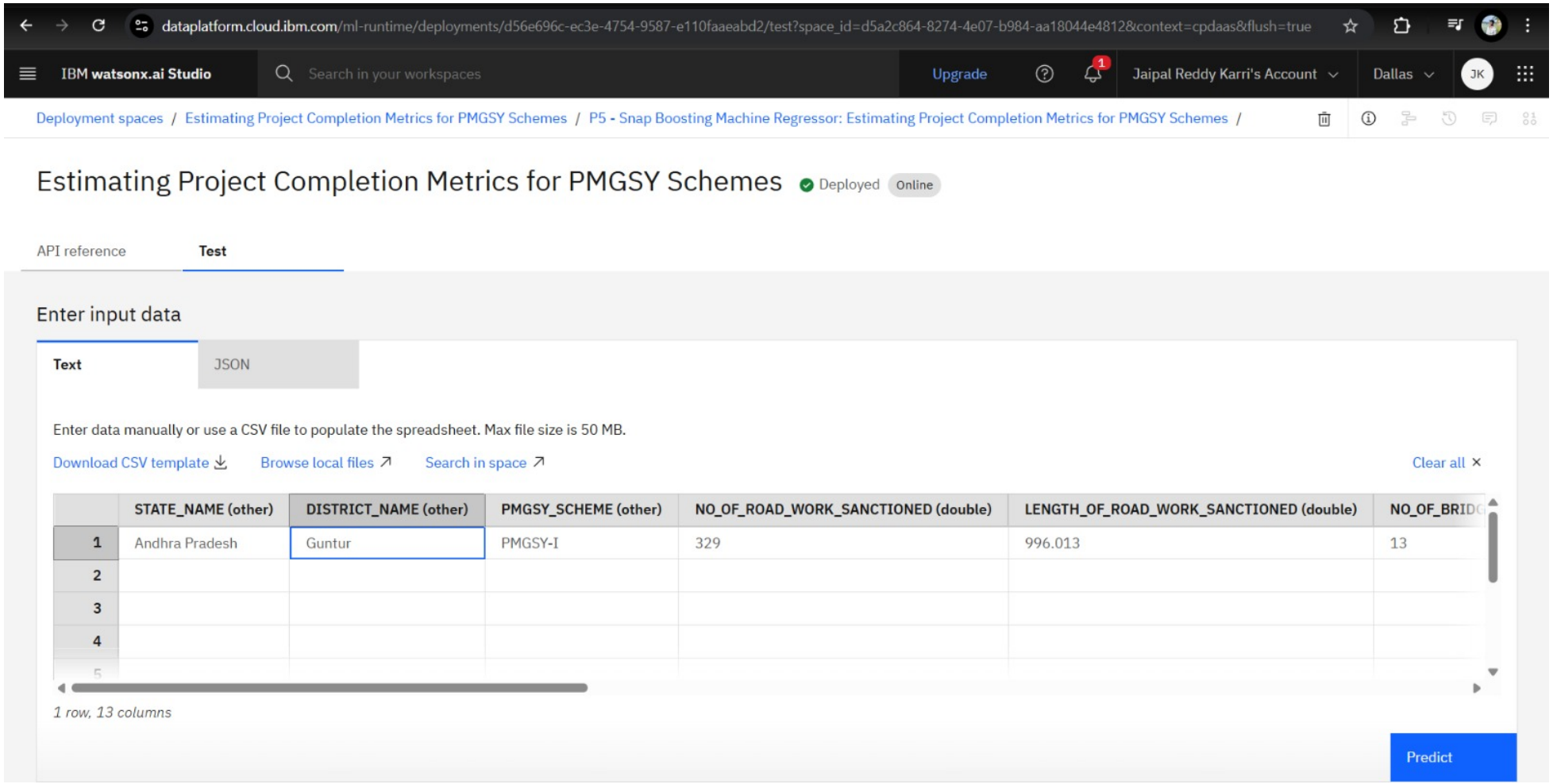
- This project uses **IBM Watsonx.ai AutoAI**, which automates model training and selection. It performs:
- **Automatic preprocessing**: Handles missing values, data types, and feature engineering.
- **Model generation**: Builds multiple pipelines using algorithms like:
  - Random Forest Regressor
  - 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

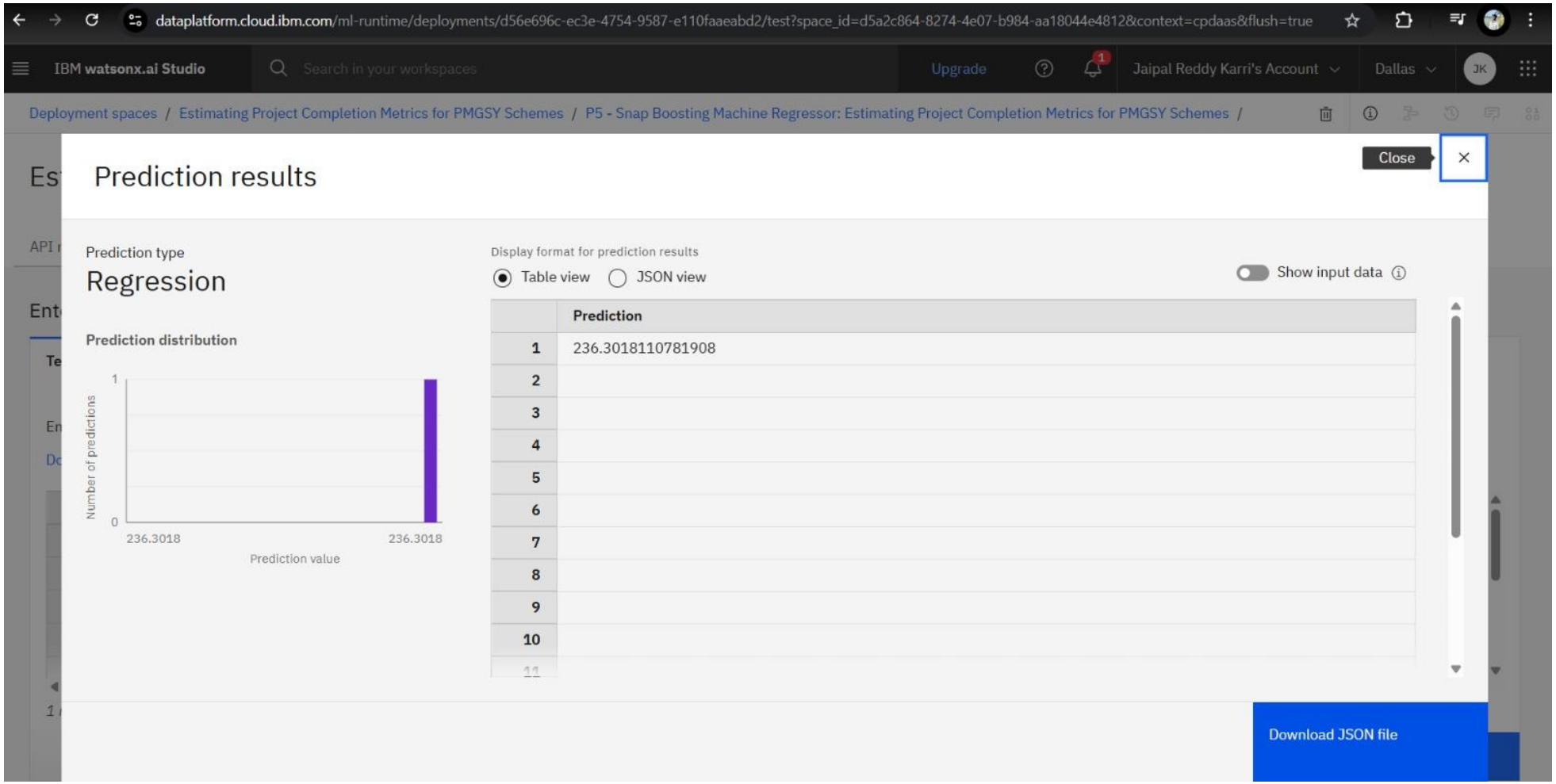
- After identifying the best-performing pipeline, the model is **deployed to the cloud** using IBM Watsonx.ai's integrated tools:
- **Model Promotion**:
  - The selected model is **promoted to a deployment space** within Watsonx.ai.
- **Deployment Creation**:
  - A **new deployment** is created (real-time 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., PMGSY scheme classification) is generated in real time.

RESULT



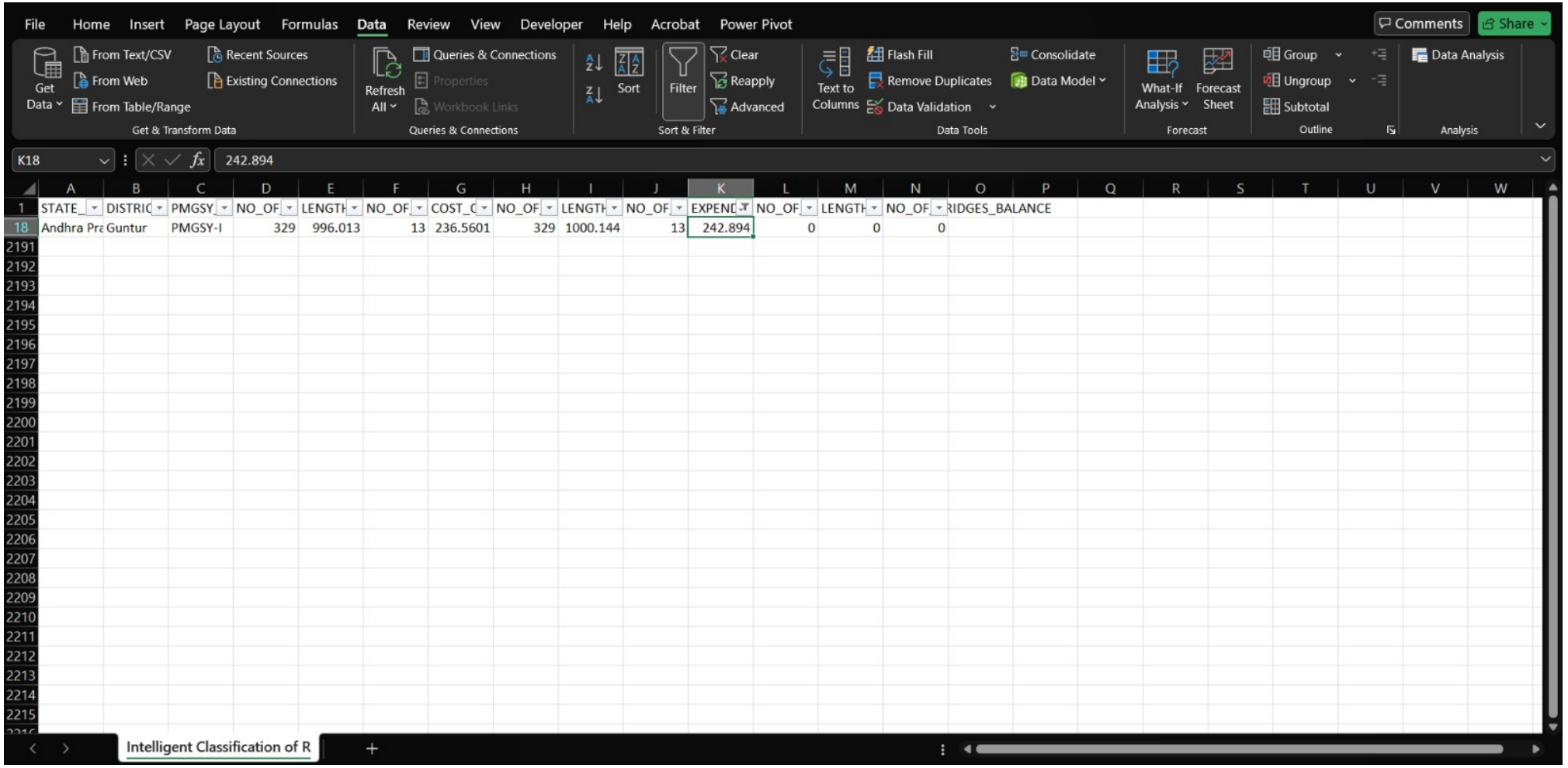
INPUT DATA

RESULT



PREDICTION RESULT

RESULT



MODEL PREDICTION ACCURACY: COMPARISON WITH ACTUAL DATA

edunet foundation

CONCLUSION

- This project demonstrated the effective use of AutoAI to build a predictive model for analyzing maternal health indicators and estimating factors influencing maternal mortality.
  By automating data preprocessing, model selection, and evaluation, the system achieved high accuracy with minimal manual intervention.
  The approach highlights the potential of AI in addressing critical public health challenges.
- The insights generated from the model reveal patterns in healthcare access, antenatal care coverage, and skilled birth attendance across different regions.
  Such findings can guide policymakers in identifying high-risk areas and prioritizing interventions.
  This ensures targeted resource allocation and supports the achievement of SDG 3.1 goals.
- Looking ahead, the system can be enhanced with real-time data integration and additional socio-economic indicators.
  Incorporating factors like literacy rates and digital accessibility can further refine predictions.
  These advancements will strengthen planning, policy formulation, and equitable healthcare delivery.

FUTURE SCOPE

- **1. Expanded Dataset Integration**
  Combine with satellite imagery or geospatial data to improve classification accuracy.
  This will enhance model robustness for projects with limited recorded details.
- **2. Real-Time Data Updates**
  Integrate live project data feeds from government portals.
  This allows automatic classification without manual intervention.
- **3. Multi-Class Hierarchical Models**
  Adopt advanced models that classify projects at both scheme and sub-scheme levels.
- **4. Predictive Policy Insights**
  Use classification outputs to forecast resource needs for upcoming projects.
  Helps in proactive budget and manpower planning.
- **5. Mobile-Friendly Platform**
  Develop a mobile UI for field officers to classify projects on-site.
  Improves accessibility and speed of updates from remote areas.

REFERENCES

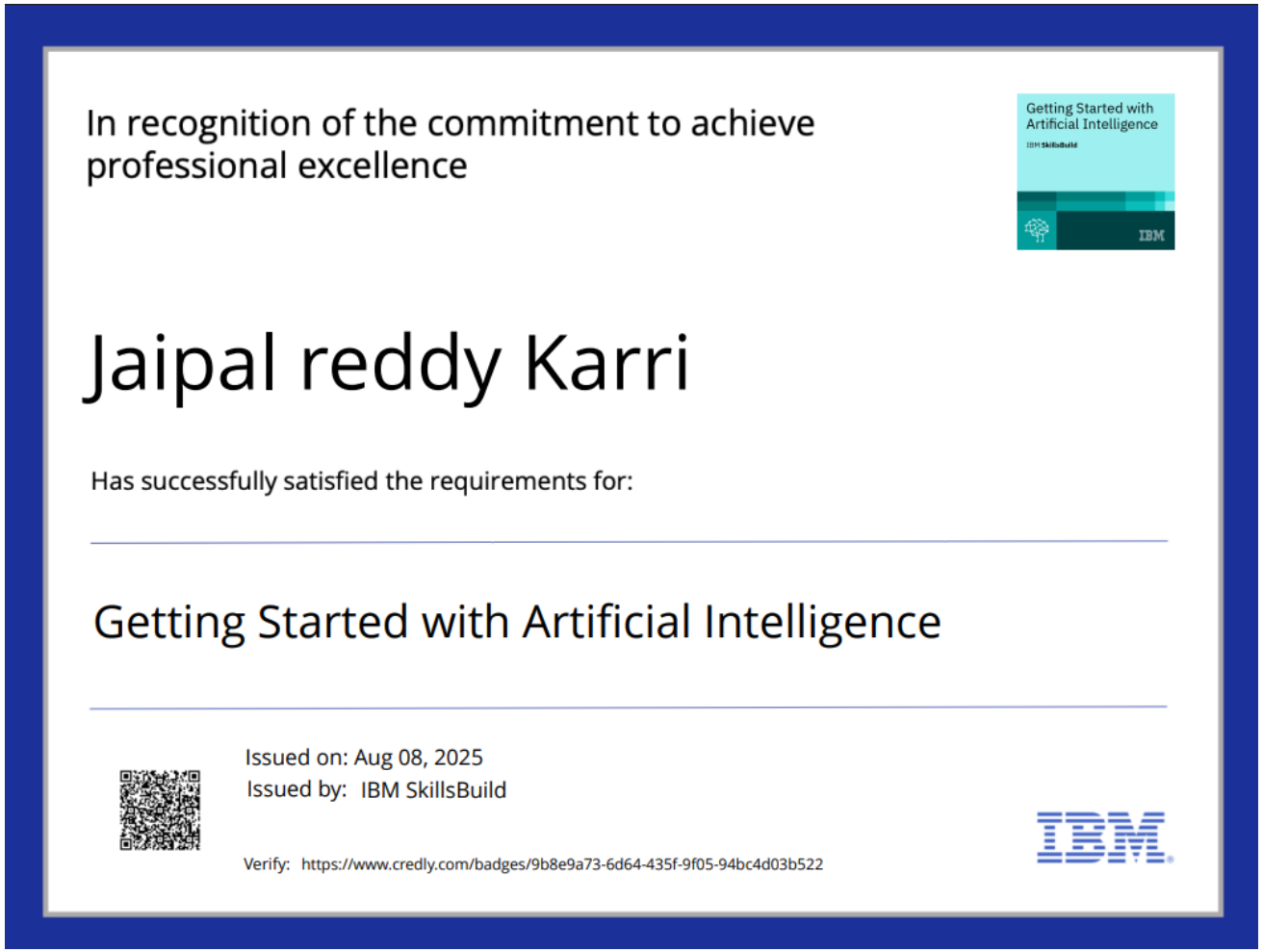
- Official Dataset Source: https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html
- IBM Cloud Platform: https://cloud.ibm.com/
- IBM Cloud: *Watsonx.ai Studio (AutoAI Tool Documentation)*

GITHUB LINK

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

edunet foundation

IBM CERTIFICATIONS



edunet foundation

IBM CERTIFICATIONS

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Jaipal reddy Karri

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

THANK YOU

edunet foundation