CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS.

Presented By:
1. Alok Anil Aher - Dr. D. Y. Patil Institute Of Technology - Artificial Intelligence & Data Science

edunet foundation

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach** (IBM Cloud, Watsonx.ai Studio)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a flagship rural development program in India, initiated to provide all-weather road connectivity to eligible unconnected habitations. Over the years, the program has evolved through different phases or schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.), each with potentially distinct objectives, funding mechanisms, and project specifications. For government bodies, infrastructure planners, and policy analysts, efficiently categorizing thousands of ongoing and completed projects is crucial for effective monitoring, transparent budget allocation, and assessing the long-term impact of these schemes. Manual classification is time-consuming, prone to errors, and scales poorly.

PROPOSED SOLUTION

- To automate the classification of rural infrastructure projects under various PMGSY schemes (PMGSY-I, PMGSY-II, PMGSY-III, etc.),
- We propose a **supervised machine learning model** trained on physical and financial characteristics of road and bridge projects.
- This solution will:
- Increase classification speed and accuracy.
- Eliminate manual errors and subjectivity.
- Enable real-time monitoring and analysis.
- The solution is implemented using IBM Watsonx.ai Studio.

SYSTEM APPROACH

The project follows a **cloud-based machine learning development lifecycle** using IBM's **Watsonx.ai Studio**, structured as follows:

- **Data Collection & Upload**
  The PMGSY dataset was sourced from AI Kosh and uploaded to IBM Cloud for processing.
- **Data Preprocessing**
  Handled missing values, encoded categorical features (like State and District), and normalized numerical columns such as road length and cost.
- **Model Development in Watsonx.ai**
  Used Watsonx.ai's **AutoAI** to automatically select and train the best classification model (e.g., Random Forest or Gradient Boosting) based on performance metrics.
- **Model Evaluation**
  Evaluated model performance using metrics like **accuracy**, **precision**, and **F1-score** to ensure reliability.
- **Model Deployment**
  The trained model was deployed as a **web service endpoint** in IBM Cloud, allowing real-time predictions through API calls.

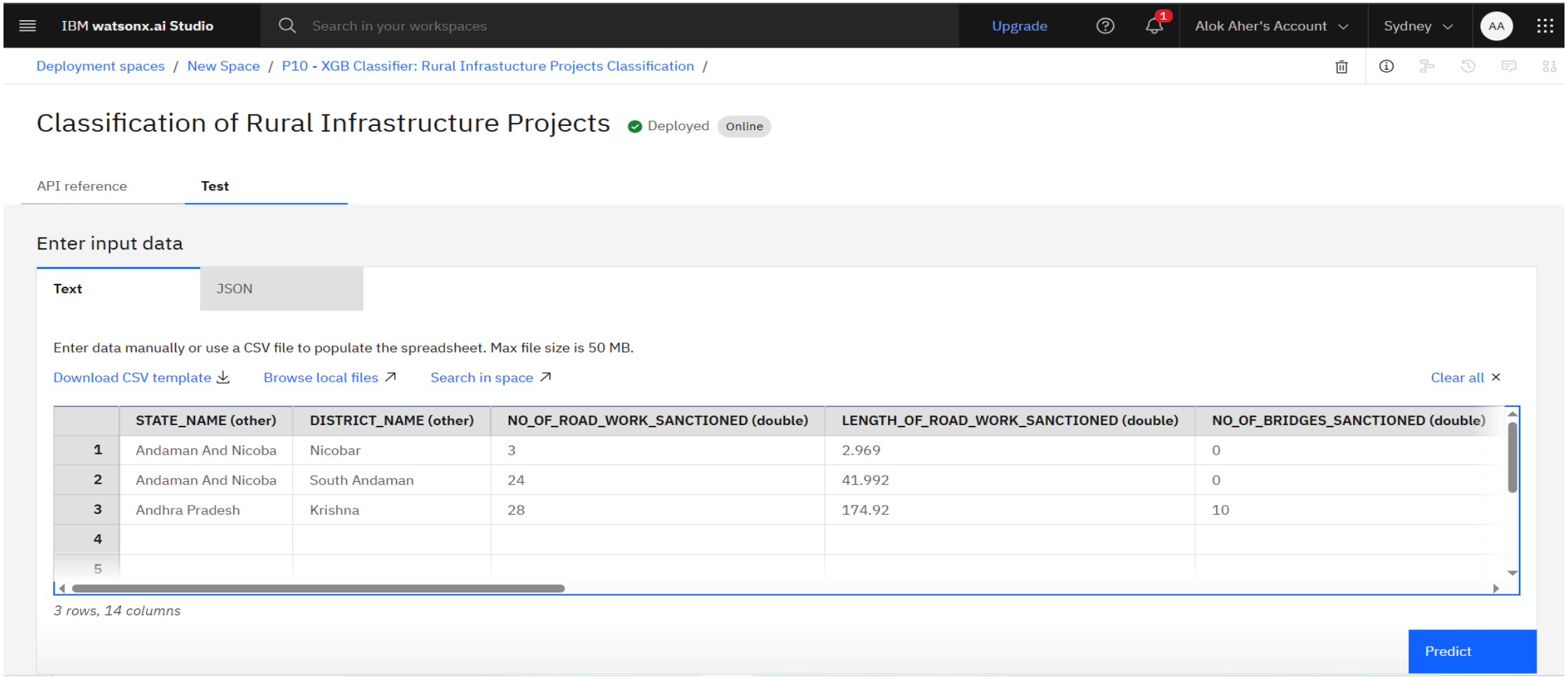
ALGORITHM & DEPLOYMENT

We used IBM Watsonx.ai Studio's **AutoAI**, which automatically evaluated and selected the best-performing classification algorithm, such as **Random Forest**, **XGBoost**, or **Gradient Boosting**, suitable for structured tabular data.

- **Input Features:**
- State Name
- District Name
- Number and Length of Road Works Sanctioned
- Number of Bridges Sanctioned
- Cost of Works Sanctioned
- Completion and Balance Details
- **Training Process:**
- Data preprocessing and feature encoding
- Dataset split into training and test sets
- AutoAI handled model training, cross-validation, and hyperparameter tuning
- The model was evaluated using metrics like **accuracy**, **precision**, and **F1-score**

- **Prediction Process:**
  The trained model predicts the appropriate PMGSY_SCHEME for new road or bridge projects based on financial and physical parameters.
- **Deployment:**
  The final model was deployed as a REST API endpoint on IBM Cloud via Watsonx.ai Studio, enabling real-time project classification through cloud-based applications or user interfaces.

RESULT :-



edunet foundation

Deployment spaces / New Space / P10 - XGB Classifier: Rural Infrastucture Projects Classification /

Classification of Rural Infrastructure Projects

Deployed Online

API reference | **Test**

Enter input data

Text | JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template | Browse local files | Search in space | Clear all

	COST_OF_WORKS_SANCTIONED (double)	NO_OF_ROAD_WORKS_COMPLETED (double)	LENGTH_OF_ROAD_WORK_COMPLETED (double)	NO_OF_BRIDGES_COMPLETED (double)	EXPENDI
1	1.3857	3	2.969	0	1.1082
2	23.42439	7	17.747	0	8.6343
3	09	11	80.703	0	49.6143
4					
5					

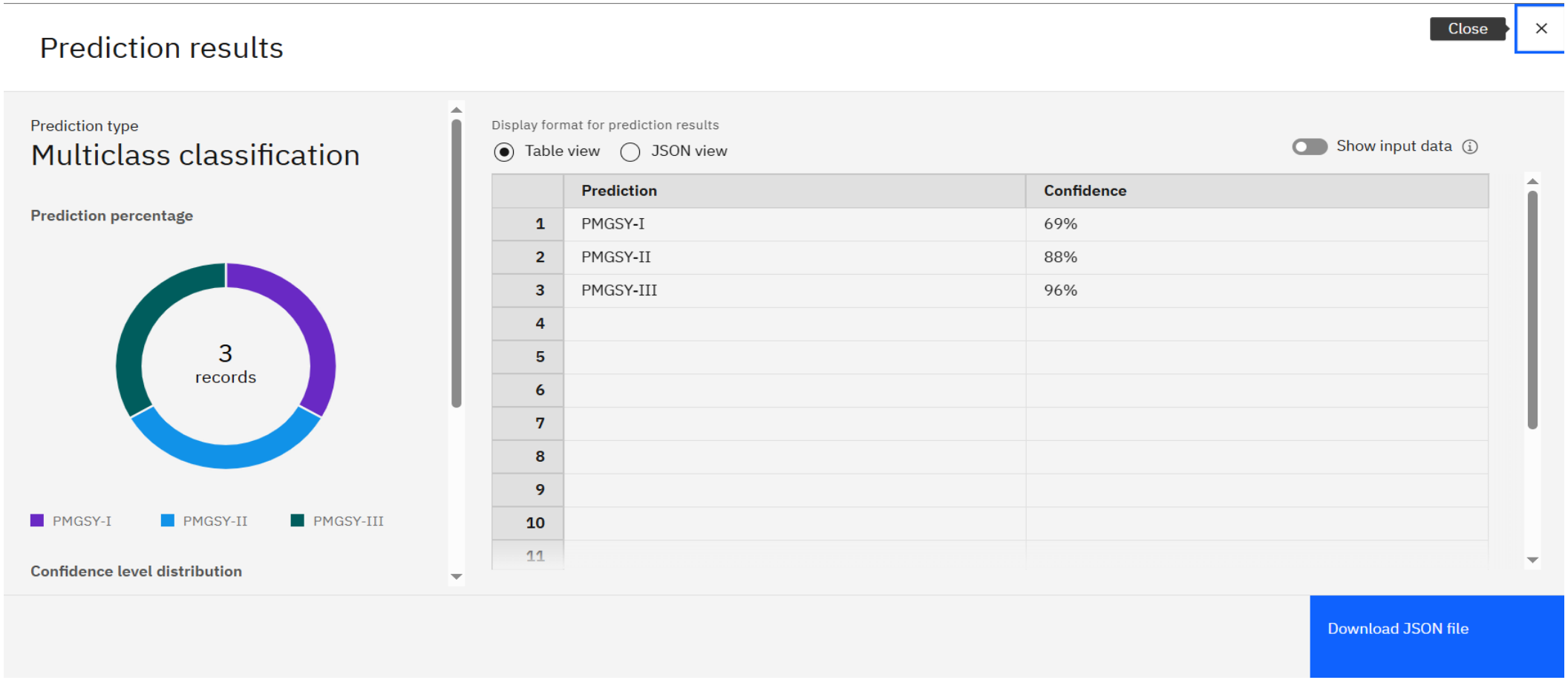
3 rows, 14 columns

Predict

Expected output from the dataset:-

Andaman /	Nicobar	PMGSY-I	3	2.969	0	1.3857	3	2.969	0	1.1082	0	0	0
Andaman /	North and	PMGSY-I	32	60.169	0	24.6908	32	59.19	0	18.5754	0	0	0
Andaman /	North and	PMGSY-II	24	54.67	0	27.7646	0	0	0	0	24	54.67	0
Andaman /	South And	PMGSY-I	32	40.146	0	18.5375	31	39.287	0	11.6894	1	0.541	0
Andaman /	South And	PMGSY-II	24	41.992	0	23.42439	7	17.747	0	8.6343	17	24.033	0
Andhra Pra	Anantapur	PMGSY-I	619	2169.505	35	526.1277	619	2126.386	34	517.912	0	0	1
Andhra Pra	Anantapur	PMGSY-II	14	125.7	0	56.0711	14	124.619	0	49.0126	0	0	0
Andhra Pra	Anantapur	PMGSY-III	27	267.158	8	180.5204	24	234.902	0	111.5182	3	23.2	8
Andhra Pra	Chittoor	PMGSY-I	283	889.681	6	188.3558	283	878.711	6	185.1493	0	0	0
Andhra Pra	Chittoor	PMGSY-II	18	126.045	0	57.9668	18	121.403	0	46.136	0	0	0
Andhra Pra	Chittoor	PMGSY-III	84	629.581	8	421.6158	68	476.394	0	193.0005	16	139.86	8
Andhra Pra	East Godav	PMGSY-I	502	1402.046	20	341.4412	500	1308.782	17	322.73	2	13.7	3
Andhra Pra	East Godav	PMGSY-II	18	123.049	0	61.7925	18	116.296	0	49.7726	0	0	0
Andhra Pra	East Godav	RCPLWEA	26	340.896	10	249.8516	17	213.208	7	157.9498	9	123.276	3
Andhra Pra	East Godav	PMGSY-III	27	183.925	2	105.3391	15	93.242	1	49.8617	12	87.78	1
Andhra Pra	East Godav	PM-JANMA	11	29.48	0		0	0	0	0	11	29.48	0
Andhra Pra	Guntur	PMGSY-I	329	996.013	13	236.5601	329	1000.144	13	242.894	0	0	0
Andhra Pra	Guntur	PMGSY-II	13	141.44	1	78.0672	13	141.428	1	69.7068	0	0	0
Andhra Pra	Guntur	PMGSY-III	34	223.804	10	246.7372	17	104.817	0	67.7782	17	115.164	10
Andhra Pra	Krishna	PMGSY-I	286	689.37	9	157.6581	286	671.76	9	155.7358	0	0	0
Andhra Pra	Krishna	PMGSY-II	14	112.79	0	74.989	14	110.499	0	64.8999	0	0	0
Andhra Pra	Krishna	PMGSY-III	28	174.92	10	170.509	11	80.703	0	49.6143	17	90.375	10

Predicted output:-



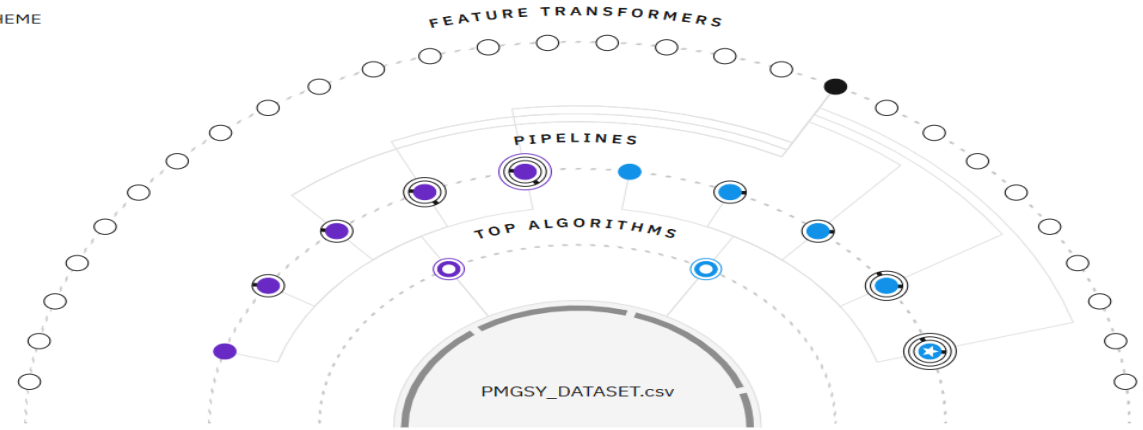
Projects / Intelligent Classification of Rural Infrastructure Projects / Rural Infrastucture Projects Classification

Experiment summary | Pipeline comparison

★ **Rank by:** Accuracy (Optimized) | Cross validation score

Relationship map

Prediction column: PMGSY_SCHEME



Progress map

Swap view

Experiment completed

10 PIPELINES GENERATED

10 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 4 minutes

View log | Save code

	Rank ↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 10**	Batched Tree Ensemble Classifier (XGB Classifier)	INCR	0.924	HPO-1 FE HPO-2 BATCH	00:01:53
	2	**Pipeline 9**	XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:49
	3	**Pipeline 8**	XGB Classifier		0.924	HPO-1 FE	00:01:11

CONCLUSION

- This project demonstrated an effective machine learning approach to classify rural infrastructure projects into their respective PMGSY schemes using project data. The model, built and deployed on **Watsonx.ai Studio**, achieved high accuracy and improved classification efficiency over manual methods.
- **Challenges** included handling missing values and high-dimensional categorical data. Potential improvements include adding geospatial features and building a user-friendly dashboard.
- Accurate classification is essential for **transparent budgeting**, **effective monitoring**, and better **policy decision-making** in rural development.

FUTURE SCOPE

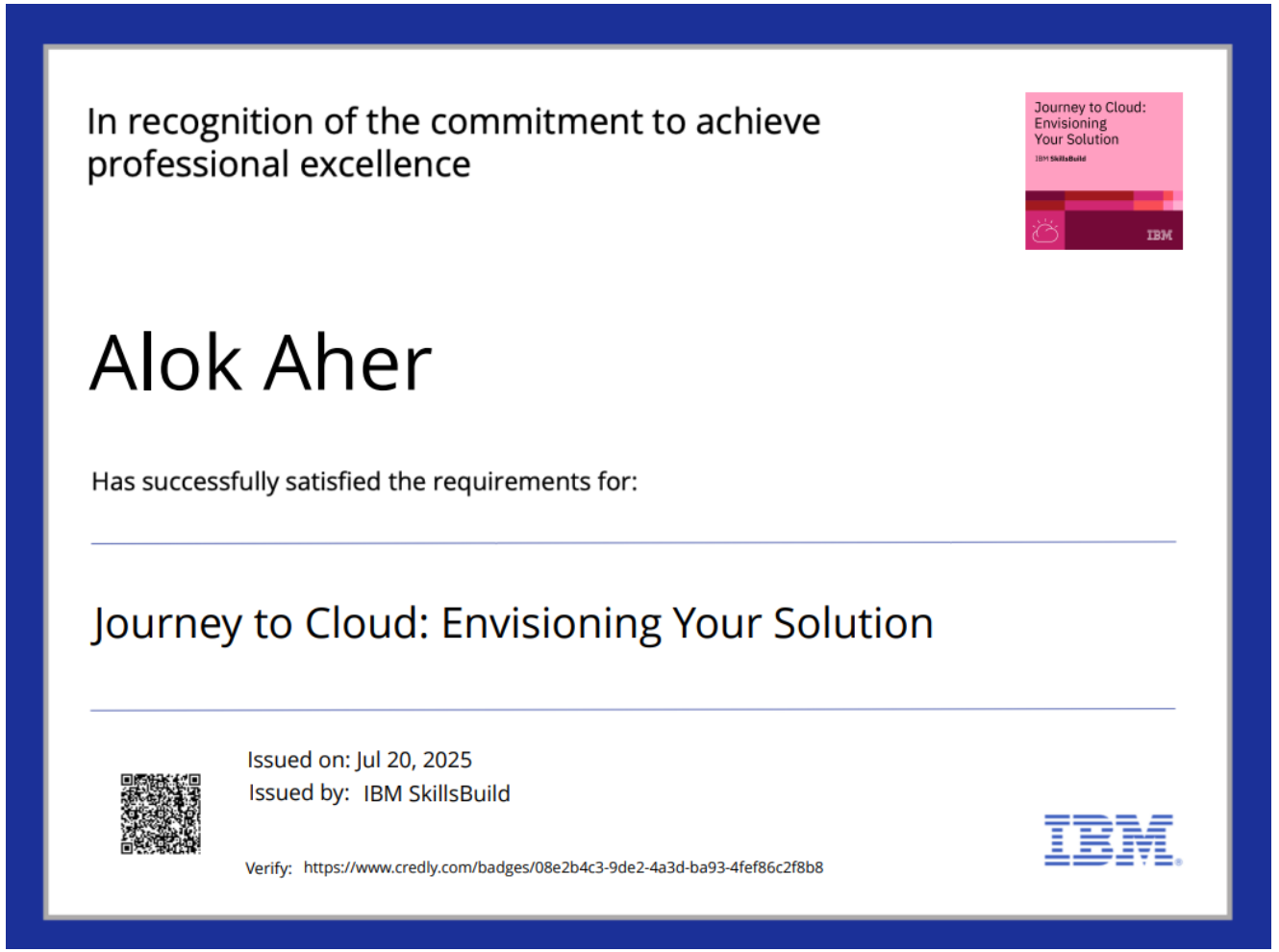
- Use advanced models like XGBoost or CatBoost for better accuracy.Add geospatial and temporal data (e.g., terrain type, project year) to enhance predictions.Build a dashboard or mobile app for easy use by government officials.
- Enable real-time feedback to improve the model continuously.Extend the model to classify projects under other rural schemes.
- Add Explainable AI features to increase transparency and trust.

REFERENCES

- https://www.ibm.com/cloud/watsonx-ai
- AI Kosh dataset link:- https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

IBM CERTIFICATIONS

IBM CERTIFICATIONS



IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Alok Aher

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