CAPSTONE PROJECT

P.S.41: INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
Megha S Bhat- S.D.M. College Of Engineering & Technology-AIML

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

The Pradhan Mantri Gram Sadak Yojana (PMGSY) aims to provide all-weather road connectivity to rural India. Over the years, the program has evolved through multiple schemes like PMGSY-I, PMGSY-II, and RCPLWEA, each with distinct objectives and funding models. Efficient classification of thousands of rural infrastructure projects into their correct schemes is essential for transparent budgeting, policy analysis, and planning. Manual classification is time-consuming and error-prone which requires an intelligent, automated solution.

PROPOSED SOLUTION

- The proposed system aims to automate the classification of rural road and bridge projects into their appropriate PMGSY schemes (e.g., PMGSY-I, PMGSY-II, RCPLWEA) using a machine learning pipeline developed and deployed entirely on **IBM Cloud Lite** services. The goal is to replace manual tagging with a scalable, fast, and accurate AI model.

- **Key Components:**

- **Data Collection: AI Kosh** dataset with physical and financial features of rural infrastructure projects.
- **Data Preprocessing:** Conducted within **AutoAI**, which handles:
  - **Missing values**: Imputed automatically using built-in strategies.
  - **Categorical variables**: Encoded via AutoAI's preprocessing pipeline.
  - **Numerical features**: Scaled/standardized where necessary.
- **Model Training:** IBM **Watson Studio's AutoAI** was used to:
  - Automate algorithm selection
  - Perform hyperparameter tuning
  - Evaluate multiple pipelines based on accuracy, precision, recall
- The **best pipeline** (Batch Train Ensemble Classifier) was selected automatically.
- **Model Deployment**: The trained model was **saved and deployed directly to IBM Watson Machine Learning (WML)**.

SYSTEM APPROACH

- The system was developed entirely using IBM Cloud. The project leverages IBM Watson Studio and AutoAI to train, evaluate and deploy the classification model.

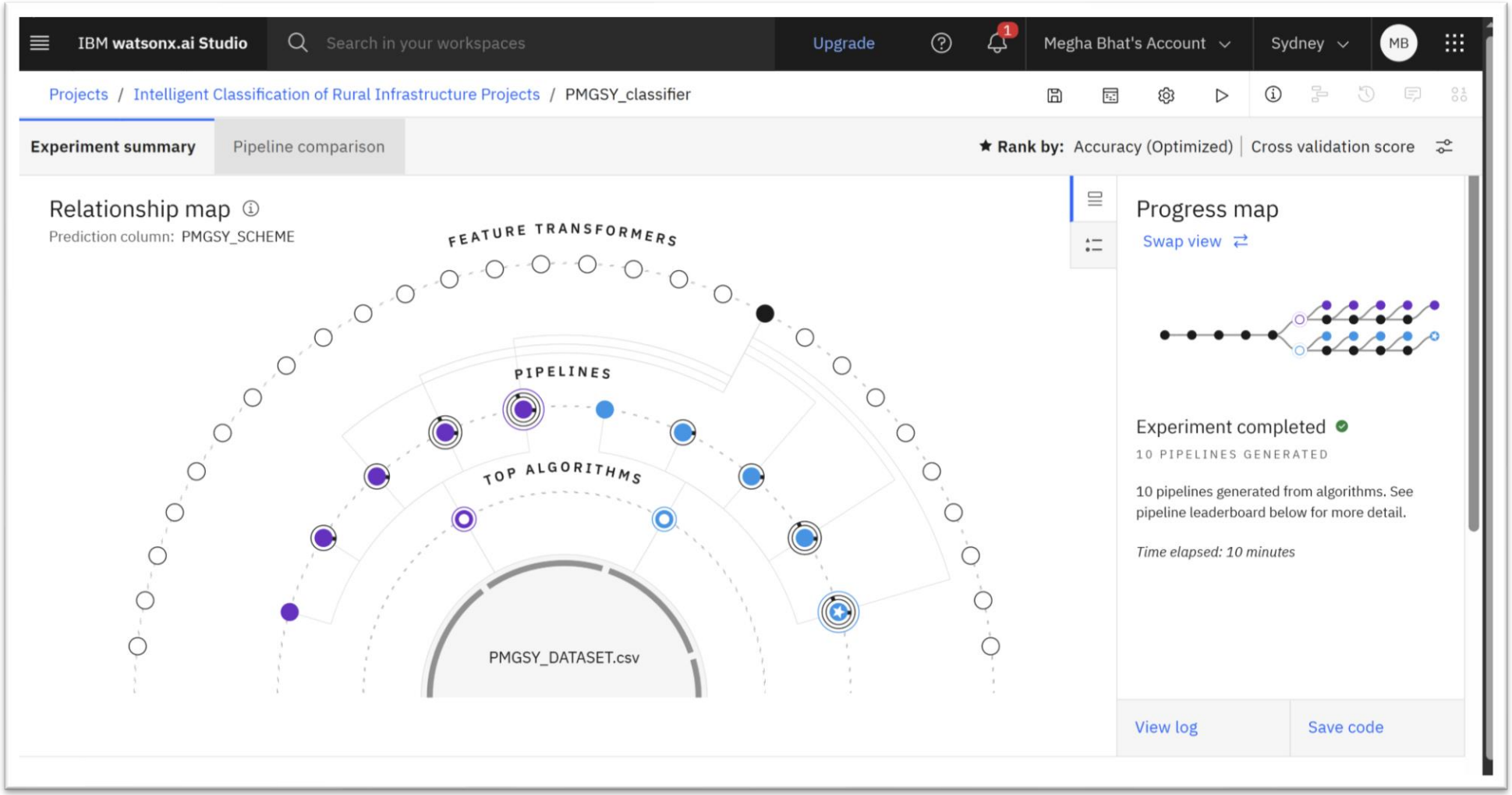
- **Technologies Used:**

- **IBM Watson Studio:** Development environment for uploading and managing data.
- **IBM Cloud Object Storage:** Used to store and access the dataset.
- **AutoAI:** This enabled automatic preprocessing, feature engineering, model selection and hyperparameter tuning.
  - Multiple models were ranked based on model accuracy and selected the best model based on highest accuracy.
- **IBM Watson Machine Learning(WML):** Used for direct model deployment.

ALGORITHM & DEPLOYMENT

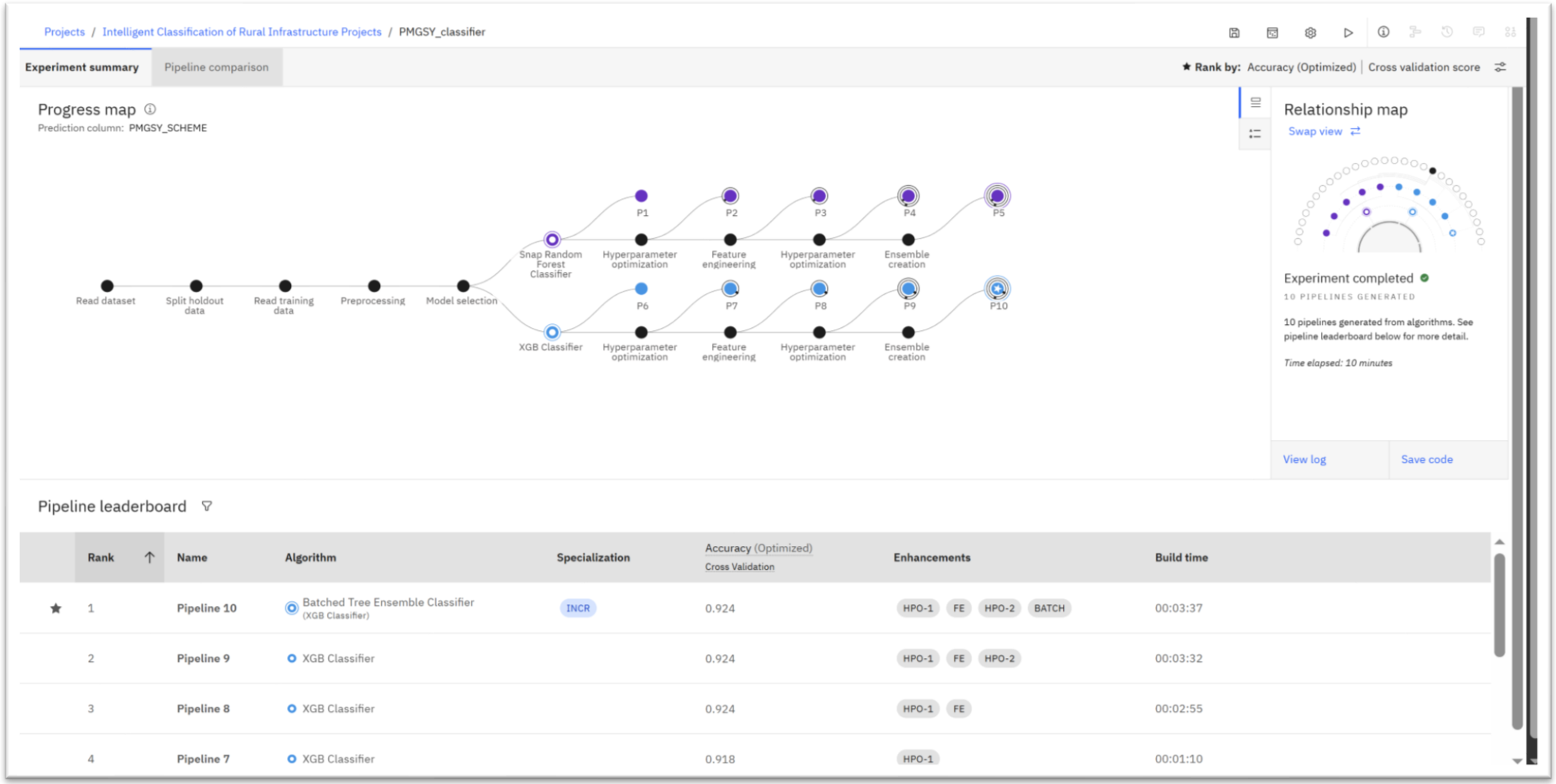
- **Algorithm Selection:**
  - **Batched Tree Ensemble Classifier** was selected based on high accuracy.
  - It combines multiple decision trees to improve prediction accuracy.
  - This algorithm is ideal for the problem because the task is a **multi-class classifiacation problem**.
  - The algorithm performed better than other models making it ideal for the classification task.
- **Data Input:**
  - **State name, district name, number and length of roads and bridges sanctioned, budget sanctioned and expenditure occurred** collected from the dataset.
  - **Target variable:** PMGSY Scheme
- **Training Process:**
  - The AutoAI handled **preprocessing of data, feature engineering** and trained multiple ML pipelines by tuning and **hyperparameter optimizations**.
  - The Batched Tree Ensemble Classifier ranked top on the leaderboard based on its accuracy in predicting the correct PMGSY scheme.
  - The model is saved and deployed on IBM Cloud.
- **Prediction Process:**
  - The selected model is deployed directly on IBM Watson Machine Learning which provides an endpoint for the model.
  - Users can enter new input data for classification. The model returns the correct **PMGSY scheme** with the confidence scores.
  - The **API key and endpoint** are used for real time predictions which can be integrated into larger applications and dashboards for infrastructure monitoring and planning.

RESULT



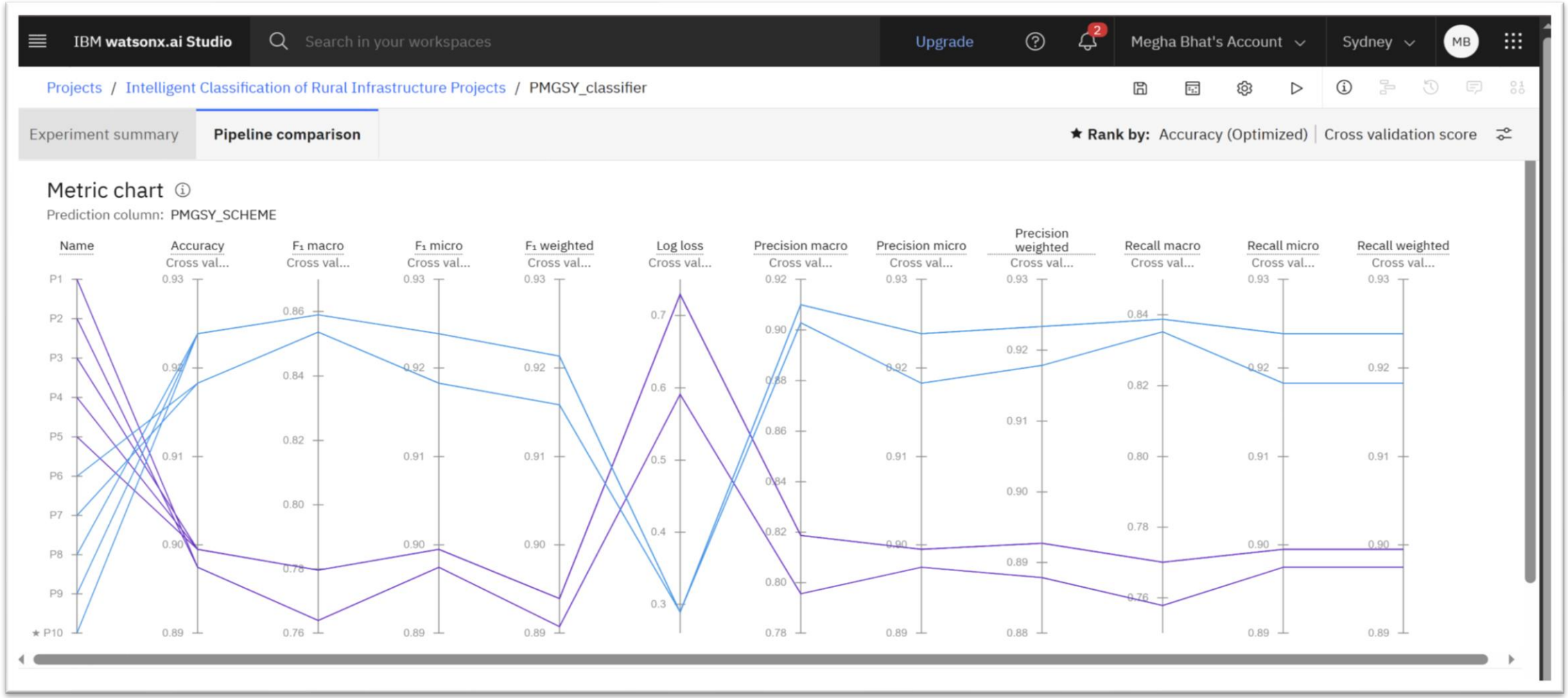
edunet foundation

RESULT



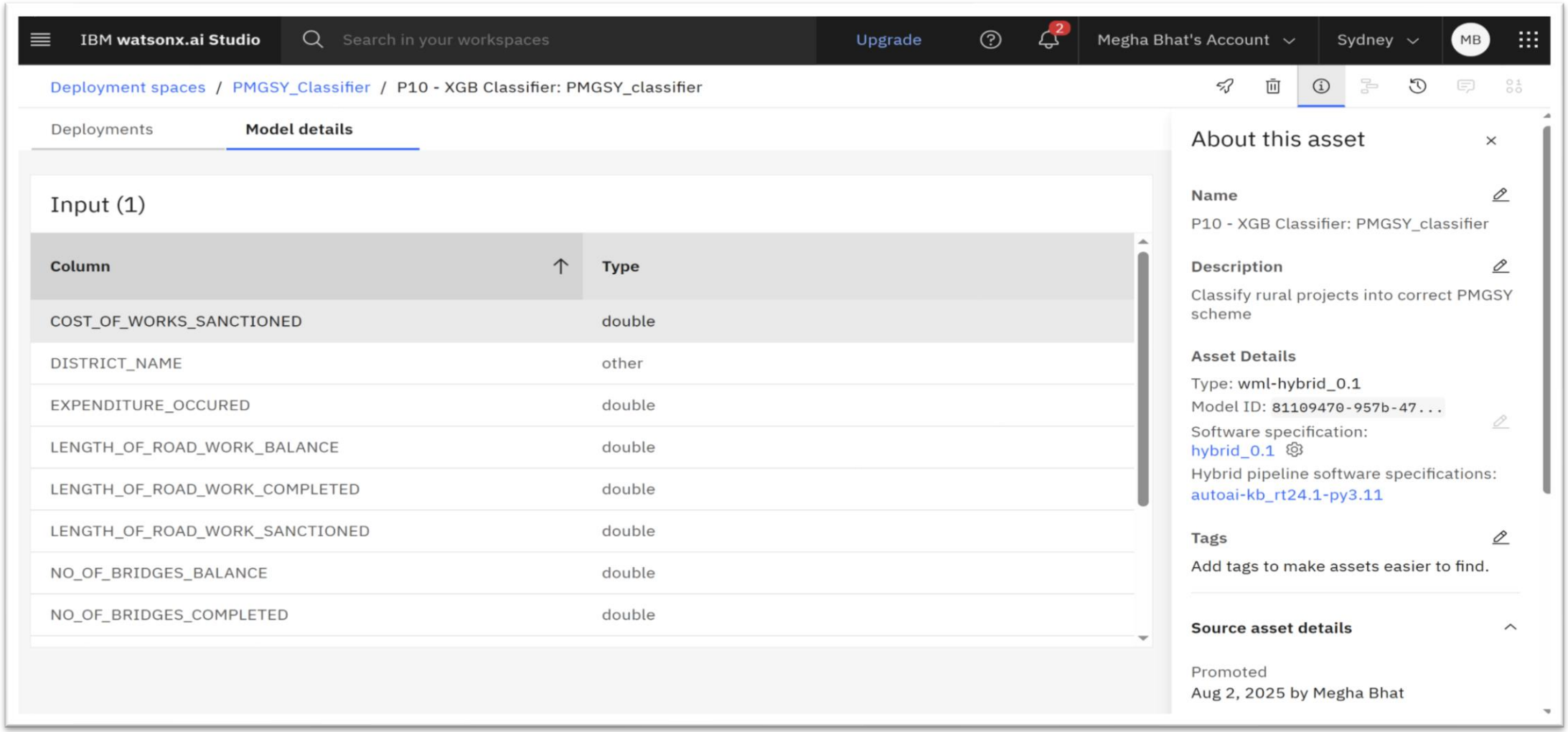
edunet foundation

RESULT

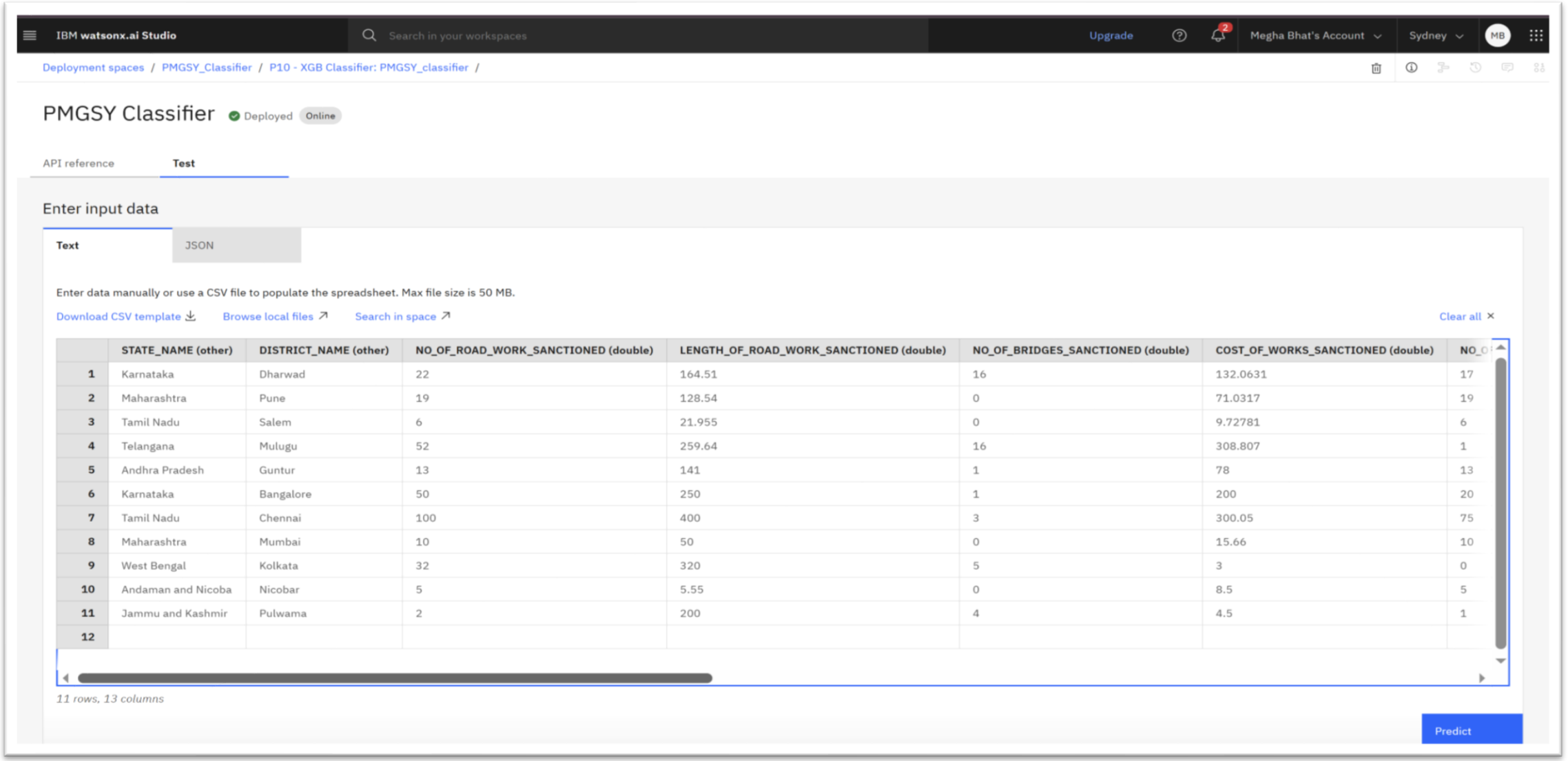


edunet foundation

RESULT

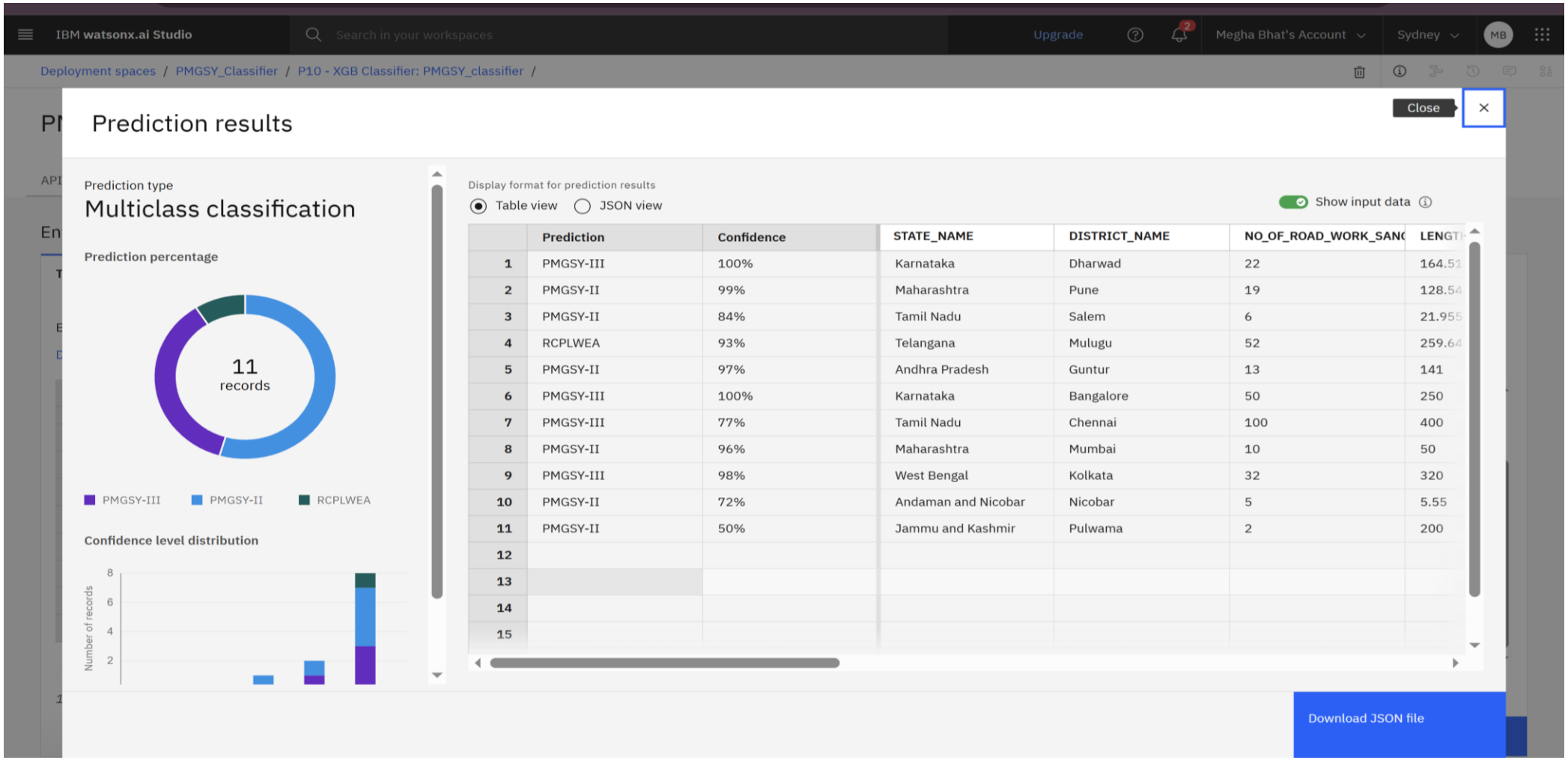


RESULT



	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF_BRIDGES_SANCTIONED (double)	COST_OF_WORKS_SANCTIONED (double)	NO_O
1	Karnataka	Dharwad	22	164.51	16	132.0631	17
2	Maharashtra	Pune	19	128.54	0	71.0317	19
3	Tamil Nadu	Salem	6	21.955	0	9.72781	6
4	Telangana	Mulugu	52	259.64	16	308.807	1
5	Andhra Pradesh	Guntur	13	141	1	78	13
6	Karnataka	Bangalore	50	250	1	200	20
7	Tamil Nadu	Chennai	100	400	3	300.05	75
8	Maharashtra	Mumbai	10	50	0	15.66	10
9	West Bengal	Kolkata	32	320	5	3	0
10	Andaman and Nicoba	Nicobar	5	5.55	0	8.5	5
11	Jammu and Kashmir	Pulwama	2	200	4	4.5	1
12							

RESULT



	Prediction	Confidence	STATE_NAME	DISTRICT_NAME	NO_OF_ROAD_WORK_SANC	LENGTH
1	PMGSY-III	100%	Karnataka	Dharwad	22	164.51
2	PMGSY-II	99%	Maharashtra	Pune	19	128.54
3	PMGSY-II	84%	Tamil Nadu	Salem	6	21.955
4	RCPLWEA	93%	Telangana	Mulugu	52	259.64
5	PMGSY-II	97%	Andhra Pradesh	Guntur	13	141
6	PMGSY-III	100%	Karnataka	Bangalore	50	250
7	PMGSY-III	77%	Tamil Nadu	Chennai	100	400
8	PMGSY-II	96%	Maharashtra	Mumbai	10	50
9	PMGSY-III	98%	West Bengal	Kolkata	32	320
10	PMGSY-II	72%	Andaman and Nicobar	Nicobar	5	5.55
11	PMGSY-II	50%	Jammu and Kashmir	Pulwama	2	200
12						
13						
14						
15						

CONCLUSION

- This project successfully demonstrated an AI-driven approach to classify rural infrastructure projects into their respective PMGSY schemes using IBM cloud. The solution provided can be scalable and efficient for automating large-scale classification tasks in rural infrastructure planning. It aims to reduce manual effort and improve decision-making.
- Key Findings:
  - Batch Tree Ensemble Classifier(XGB Classifier) provided high accuracy and robust performance.
  - The system handled mixed data types and multi-class classification effectively.
  - Deployment was made directly on IBM cloud without custom backend deployment.

FUTURE SCOPE

- Real-time dashboard integration for government analysts and policy planners.
- Expanding the dataset by incorporating satellite data and demographic information.
- Integration with mapping tools to visualize the project distribution.
- Multi-language support can be provided for better understanding by people across different states.
- Feedback loop can be integrated to continuously improve the model using real-world validation.

REFERENCES

[1] Government of India, “Pradhan Mantri Gram Sadak Yojana(PMGSY) Dataset”, AI Kosh, https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

[2] IBM “watsonx.ai Studio Documentation”, https://dataplatform.cloud.ibm.com/docs/content/svc-welcome/wsl.html?context=cpdaas

[3] IBM “watsonx.ai Runtime Documentation”, https://dataplatform.cloud.ibm.com/docs/content/svc-welcome/wml.html?context=cpdaas

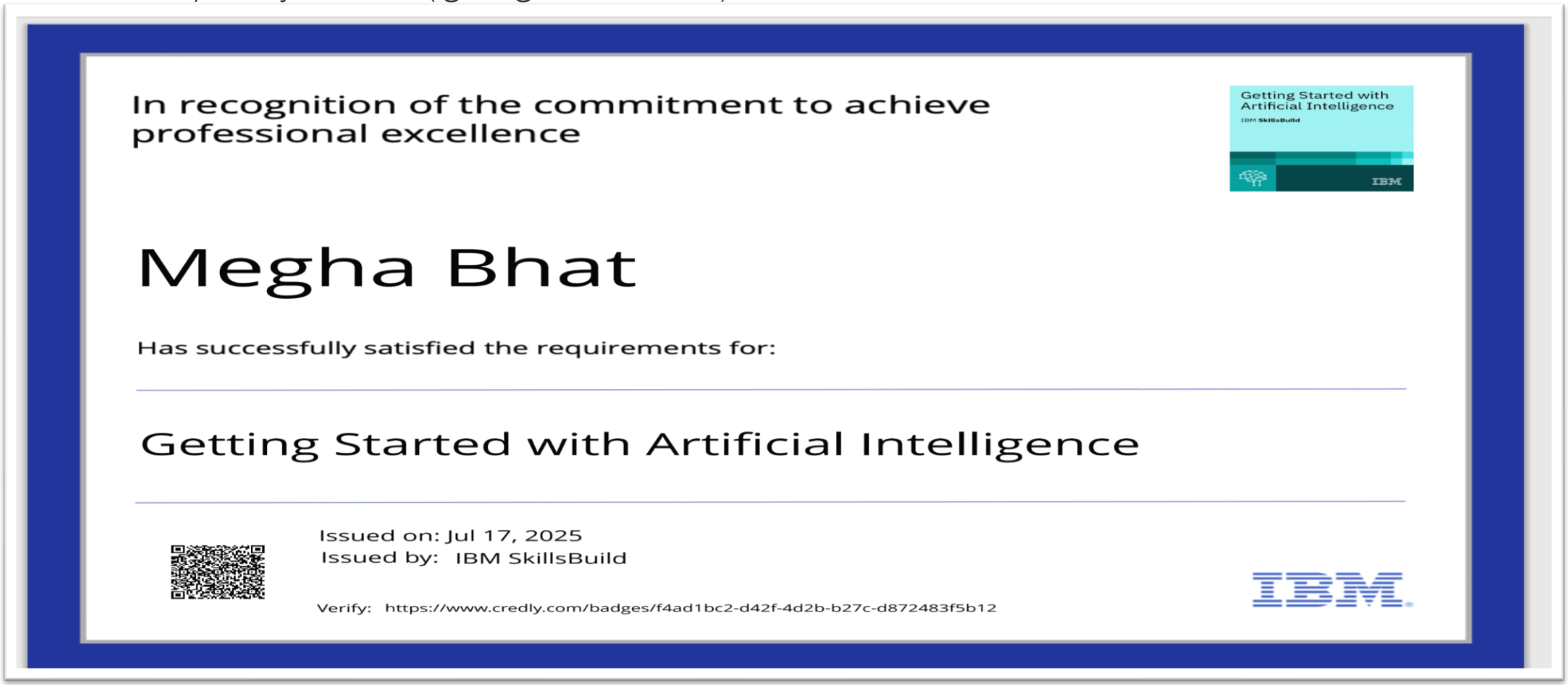
[4] “Watson Machine Learning - IBM Documentation”, https://www.ibm.com/docs/en/software-hub/5.2.x?topic=services-watson-machine-learning

[5] Government of India, “Pradhan Mantri Gram Sadak Yojana(PMGSY) Overview”, Ministry of Rural Development, https://pmgsy.nic.in

[6] “Managing User API Key Documentation”, https://cloud.ibm.com/docs/account?topic=account-userapikey&interface=ui

IBM CERTIFICATIONS

- Screenshot/ credly certificate(getting started with AI)



IBM CERTIFICATIONS

- Screenshot/ credly certificate(Journey to Cloud)

IBM CERTIFICATIONS

- Screenshot/ credly certificate(RAG Lab)

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Megha Bhat

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU

edunet foundation