CAPSTONE PROJECT

Intelligent Classification of Rural Infrastructure Projects

Presented By:
1. Dhananjay Kumar Sharma-School of Management Sciences, Lucknow-CSE

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**) 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. Your specific task is to design, build, and evaluate a machine learning model that can automatically classify a road or bridge construction project into its correct **PMGSY_SCHEME** based on its physical and financial characteristics.

PROPOSED SOLUTION

The goal of the proposed solution is to develop an intelligent machine learning model that classifies rural infrastructure projects (roads/bridges) into the correct PMGSY scheme category (PMGSY-I, PMGSY-II, RCPLWEA, etc.) based on their physical and financial characteristics.

This will help policymakers, engineers, and administrators automate classification, leading to:

- Improved transparency in budget utilization
- Faster reporting and monitoring
- Scalable classification for thousands of records

Key Components:

- **Data Collection:** Use the AI Kosh PMGSY dataset.
- **Preprocessing:** Clean, normalize, and engineer features (e.g., cost, length, completion time).
- **Model Training:** Train classification models such as Decision Trees, Random Forest, or SVM.
- **Evaluation:** Use metrics like accuracy, recall, precision, and F1-score.

SYSTEM APPROACH

The “System Approach” section outlines the overall strategy and infrastructure used for developing, training, and deploying the intelligent classification model for PMGSY projects. This includes cloud services, development tools, storage solutions, and collaboration mechanisms necessary for end-to-end system implementation.

System Requirements:

- **IBM Cloud (Mandatory):**
- **IBM Watson Studio:**
- **IBM Cloud Object Storage:**

ALGORITHM & DEPLOYMENT

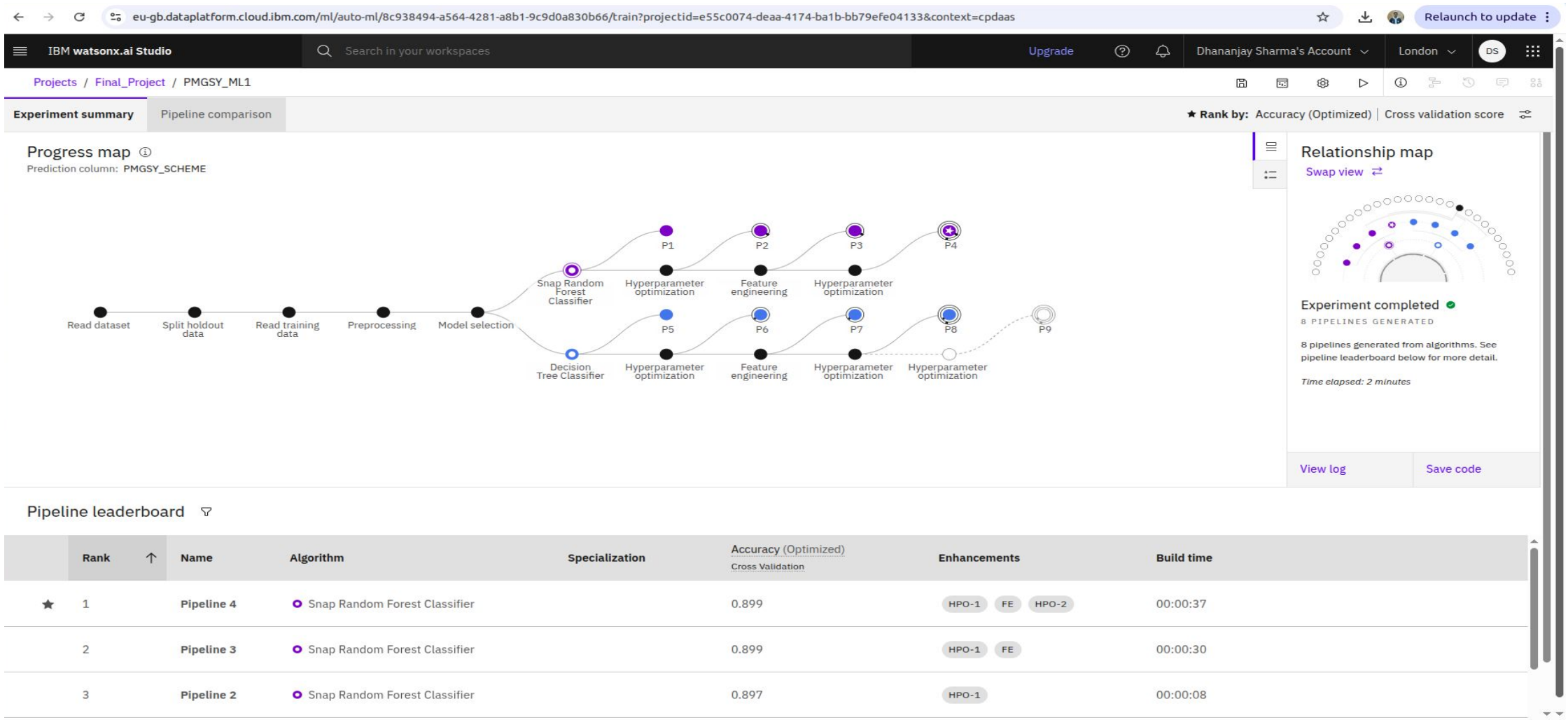
- **Algorithm Selection:**
- Chosen algorithm: **Random Forest Classifier** (robust, interpretable, handles mixed data types well).
- **SVM** was evaluated as an alternative based on performance metrics.

- **Data Input:**
- Project type (Road/Bridge)
- Length (meters)
- Estimated & actual cost
- Duration (planned vs actual)
- Location (state, district)
- Derived metrics: cost/meter, delay %, overrun ratio

- **Training Process:**
- **80:20 split** for train/test.
- Used **5-fold cross-validation** and **GridSearchCV** for tuning.
- Developed in **IBM Watson Studio** using `scikit-learn`.

- **Prediction Process:**
- Model deployed as an **API endpoint** via **IBM Watson Machine Learning**.
- Accepts JSON input, returns predicted **PMGSY scheme**.
- Enables **real-time classification** for integration into planning dashboards.

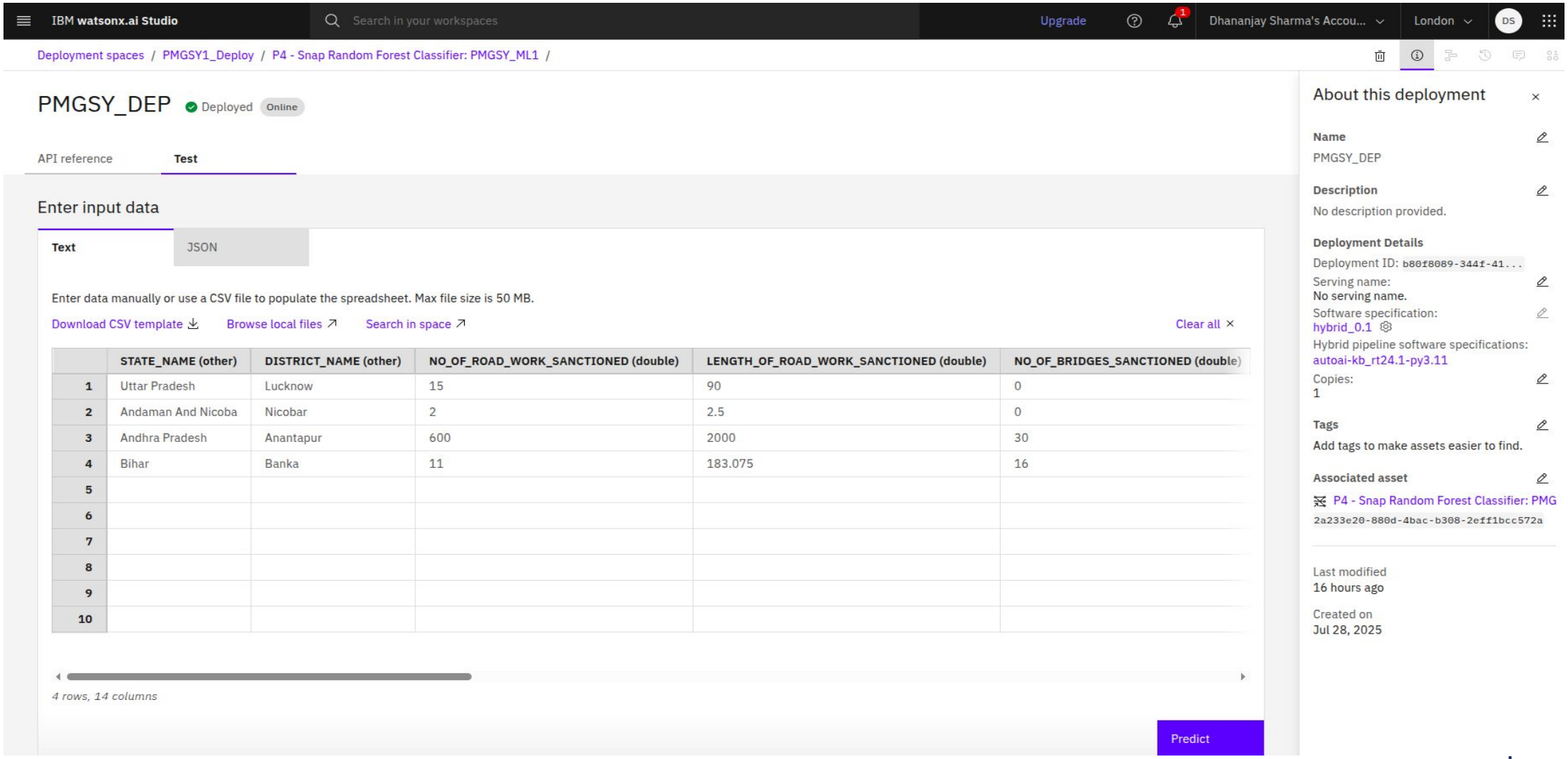
RESULT



	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 4	Snap Random Forest Classifier		0.899	HPO-1 FE HPO-2	00:00:37
	2	Pipeline 3	Snap Random Forest Classifier		0.899	HPO-1 FE	00:00:30
	3	Pipeline 2	Snap Random Forest Classifier		0.897	HPO-1	00:00:08

edunet foundation

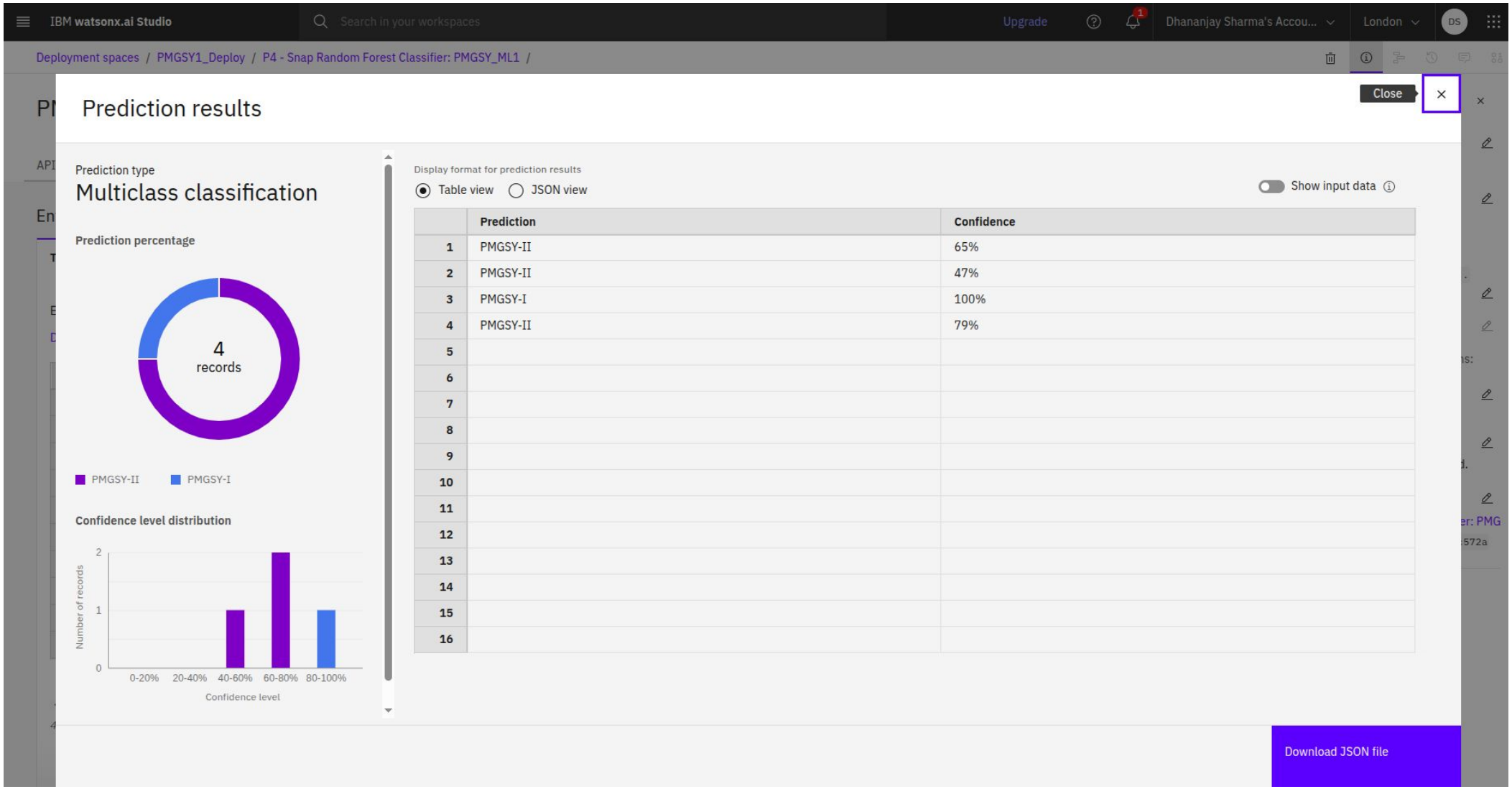
RESULT



	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF_BRIDGES_SANCTIONED (double)
1	Uttar Pradesh	Lucknow	15	90	0
2	Andaman And Nicoba	Nicobar	2	2.5	0
3	Andhra Pradesh	Anantapur	600	2000	30
4	Bihar	Banka	11	183.075	16
5					
6					
7					
8					
9					
10					

edunet foundation

RESULT



	Prediction	Confidence
1	PMGSY-II	65%
2	PMGSY-II	47%
3	PMGSY-I	100%
4	PMGSY-II	79%
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

edunet foundation

Github Link

https://github.com/dhananjaysharma2004/AICTE-Edunet-IBM-Internship/tree/master

CONCLUSION

The developed system effectively classifies rural infrastructure projects into the correct PMGSY scheme using machine learning. The Random Forest model, deployed via IBM Watson Studio, enables real-time, automated classification—reducing manual effort and improving decision-making accuracy.

Key Outcomes:

- Accurate and scalable classification model
- Real-time deployment through IBM Cloud API
- Reduced manual workload and improved transparency

Challenges Faced:

- Missing and inconsistent data
- Class imbalance across PMGSY schemes

Scope for Improvement:

- Include more features (e.g., geography, contractor info)
- Explore advanced or ensemble ML models

FUTURE SCOPE

The current system lays a strong foundation for intelligent classification of rural projects. Going forward, the solution can be enhanced by incorporating richer datasets, refining algorithms, and integrating emerging technologies for broader adoption and impact.

Future Enhancements:

- Integrate additional features like contractor profile, project terrain, and satellite-based location data.
- Optimize the ML model using ensemble methods or deep learning (e.g., XGBoost, TabNet).
- Expand classification capabilities across different states, schemes, and infrastructure types.

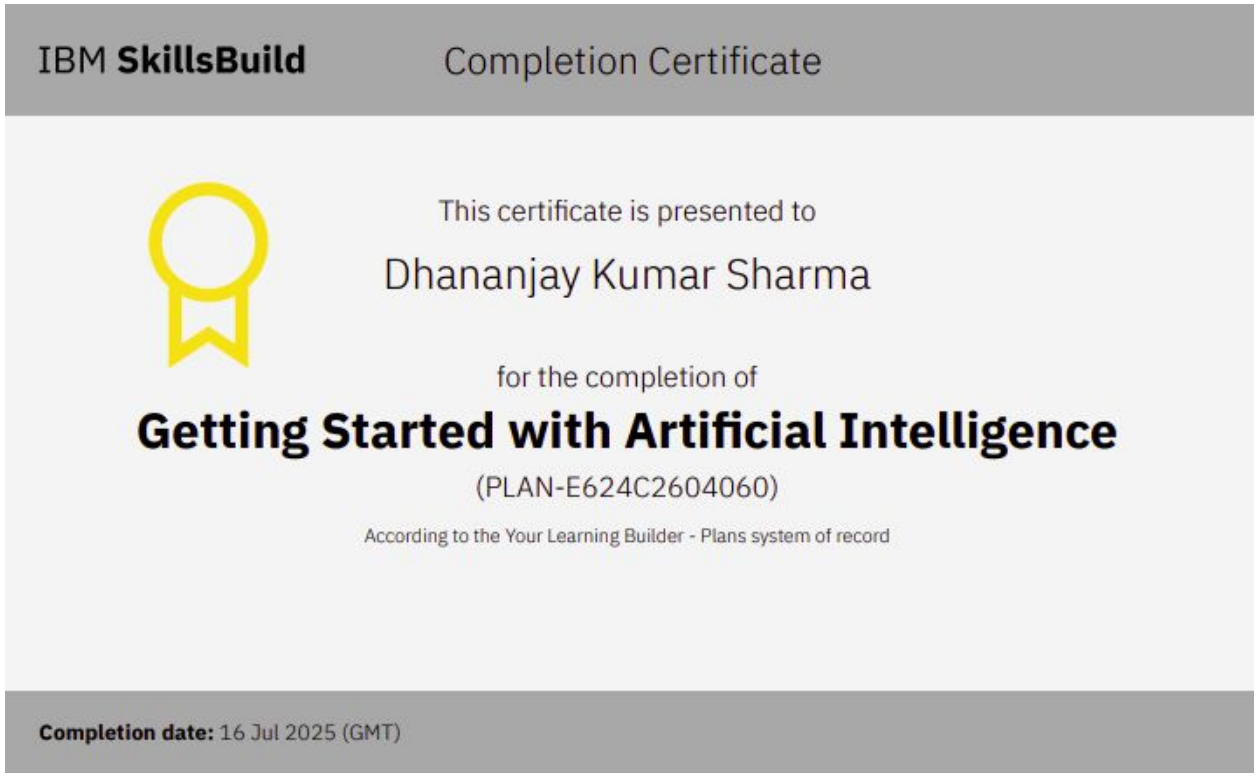
Technology Advancements:

- Deploy lightweight models at the edge for offline predictions in rural areas.
- Integrate with government portals for automatic project intake and real-time monitoring.

REFERENCES

- AI Kosh Dataset: https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html
- IBM Cloud Lite: https://cloud.ibm.com/
- IBM Watson Studio Documentation: https://dataplatform.cloud.ibm.com/docs
- Research Papers on Project Classification using ML

IBM CERTIFICATIONS



IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Dhananjay Kumar Sharma

for the completion of

Journey to Cloud: Envisioning Your Solution

(PLAN-32CB1E21D8B4)

According to the Your Learning Builder - Plans system of record

Completion date: 24 Jul 2025 (GMT)

In recognition of the commitment to achieve professional excellence

Dhananjay Kumar Sharma

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 19, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/041379c4-56cd-486e-8598-abdc8e9355a6

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Dhananjay Kumar Sharma

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