# CAPSTONE PROJECT

# INTELLIGENT CLASSIFICATION OF PMGSY PROJECTS USING MACHINE LEARNING

**Presented By:**
**Amarjit Kumar Natraj-G H Raisoni College of Engineering-CSE**

edunet foundation

# OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

# PROBLEM STATEMENT

The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a flagship rural development program aimed at providing all-weather road connectivity to unconnected habitations across India. Over the years, PMGSY has evolved into multiple schemes (PMGSY-I, PMGSY-II, RCPLWEA) with distinct objectives and specifications. Manually classifying thousands of ongoing and completed projects into these schemes is tedious, error-prone, and inefficient, making it challenging for government bodies to manage project monitoring, transparent budgeting, and impact assessment effectively.

# PROPOSED SOLUTION

The proposed system aims to automate the classification of PMGSY infrastructure projects into their correct schemes (PMGSY-I, PMGSY-II, RCPLWEA) using machine learning. The system involves:

**Data Collection:**

- Collect project data (road length, bridges, costs, expenditures) from AI Kosh.

**Data Preprocessing:**

- Clean data, handle missing values, and engineer relevant features.

**Machine Learning Algorithm:**

- Train an **XGBoost Classifier** to classify projects based on their physical and financial attributes.
- Evaluate using Accuracy Score and Confusion Matrix.

**Deployment:**

- Deploy the model as an API using **IBM Watson Machine Learning**.
- Provide a simple interface/API for real-time project classification.

**Evaluation:**

- Continuously monitor model performance and refine based on feedback.

**Result:**

- Automated, accurate, and scalable project classification system reducing manual efforts and improving decision-making efficiency.

# SYSTEM APPROACH

**The "System Approach" section outlines the overall strategy and methodology for developing and implementing the rental bike prediction system.**

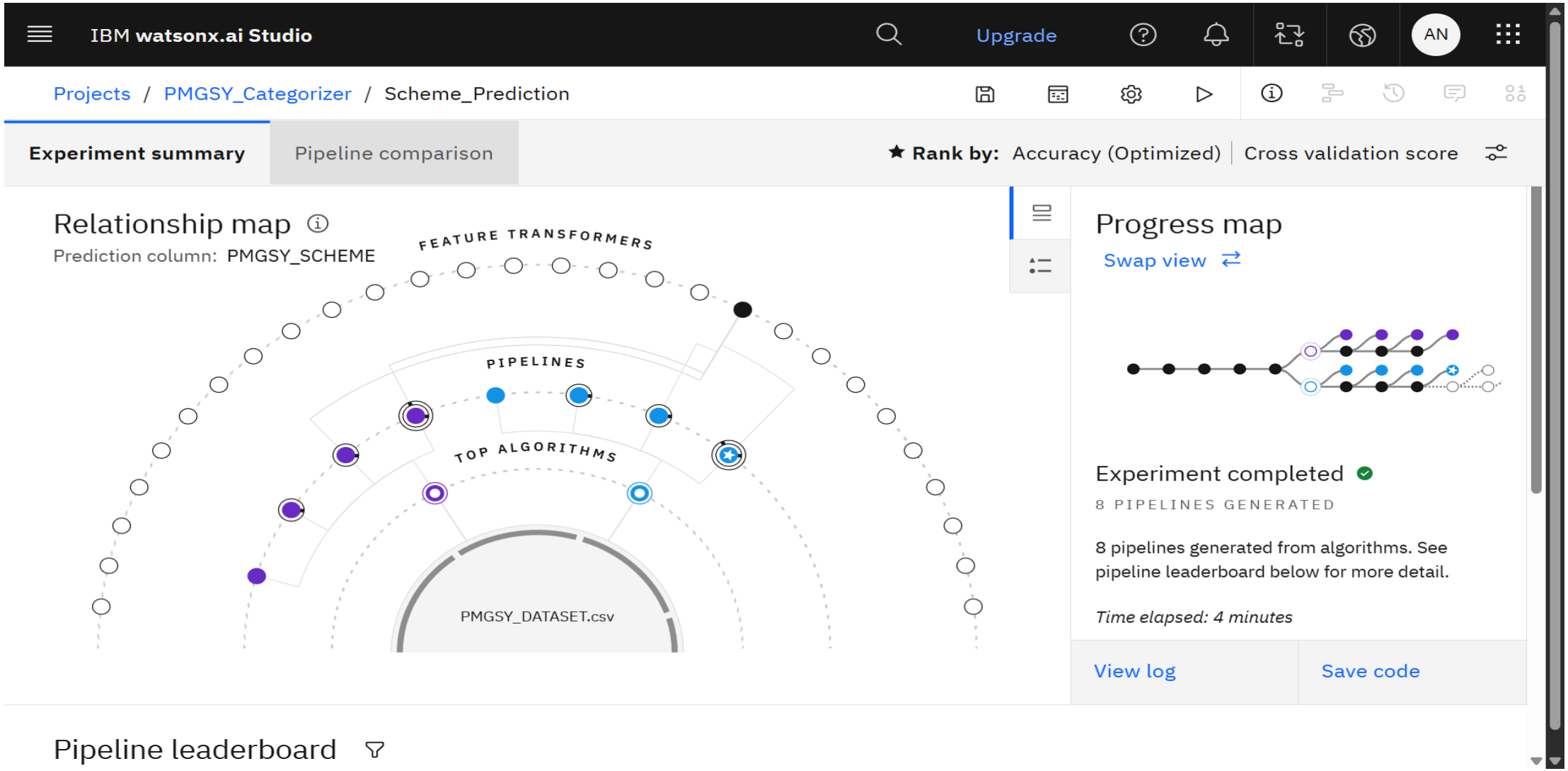
**System requirements :-**

- IBM Cloud (mandatory)
- IBM Watson Studio for model development and deployment
- IBM Cloud Object Storage for dataset handling and storage
- IBM Watson Machine Learning for API deployment

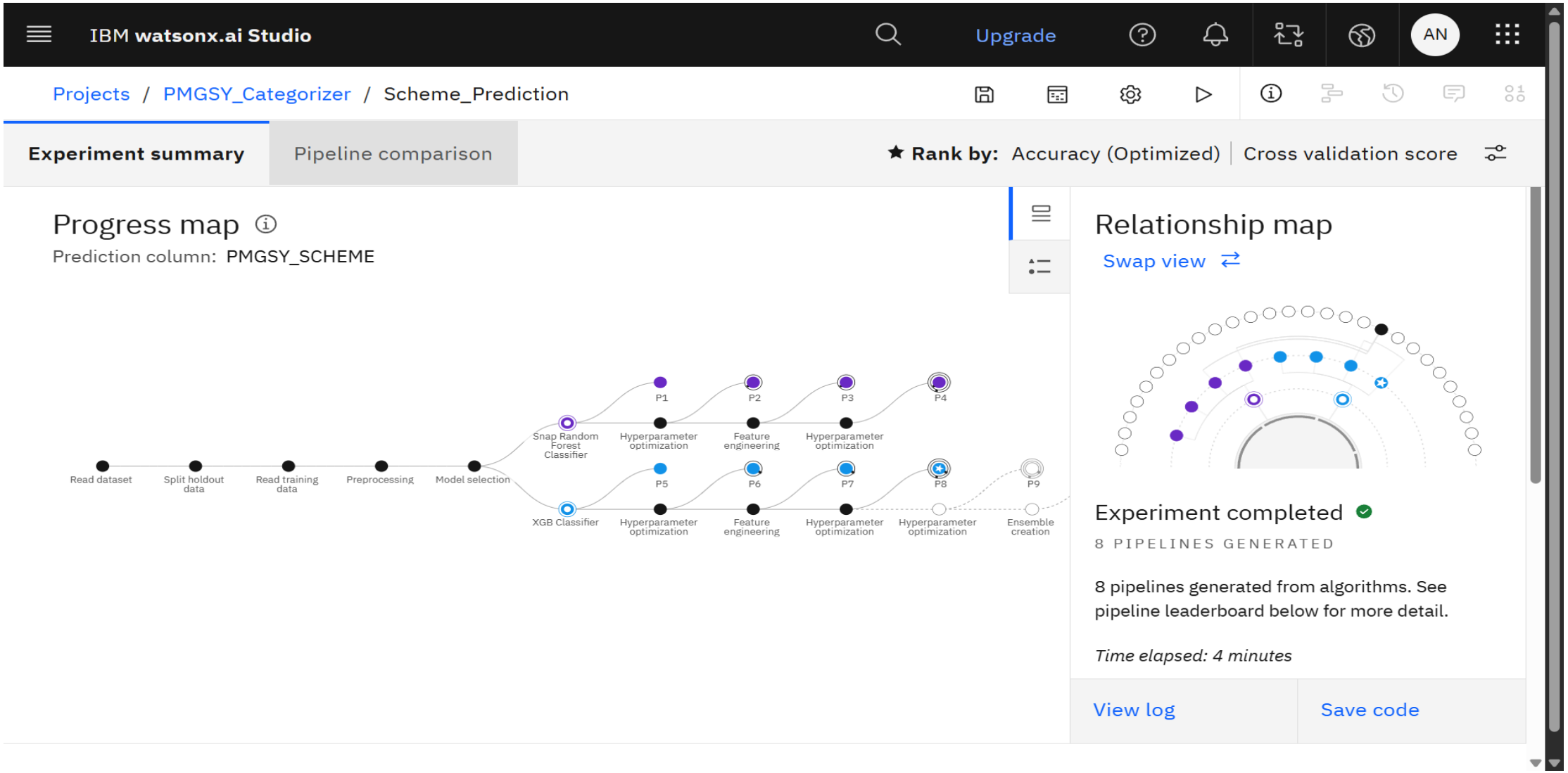
# ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  Used **XGBoost Classifier** for its high accuracy on structured data and ability to handle complex feature interactions relevant for PMGSY scheme classification.
- **Data Input:**
  Project attributes like road length, number of bridges, sanctioned costs, expenditures, and derived features (e.g., % completion).
- **Training Process:**
  Performed data cleaning, feature selection, and trained the model with an 80-20 Train-Test split. Hyperparameters were tuned to improve accuracy.
- **Prediction & Deployment:**
  Deployed the trained model as an API using **IBM Watson Machine Learning**. Users input project data, and the API returns the predicted PMGSY scheme in real-time.

# RESULT

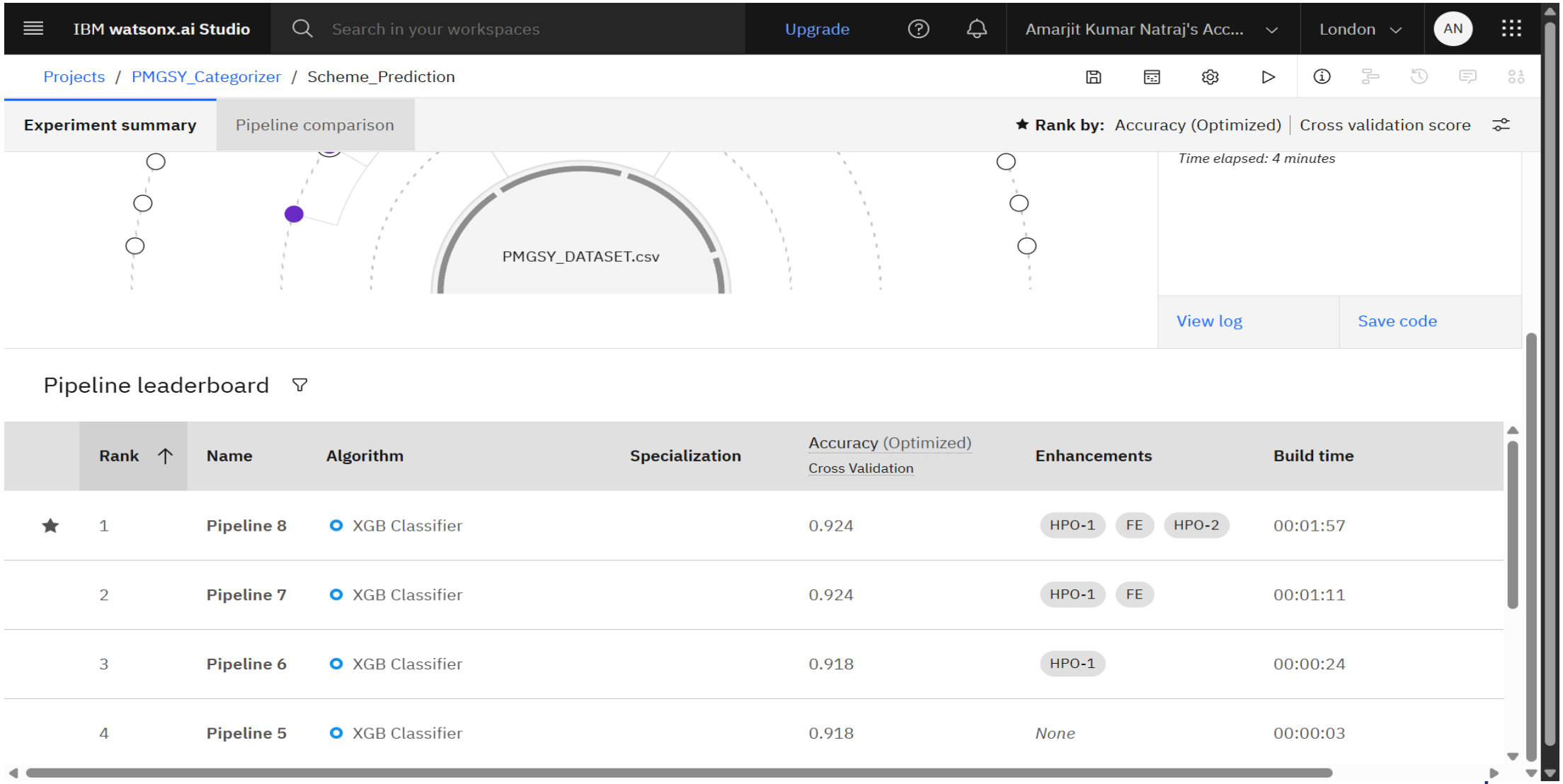


# RESULT



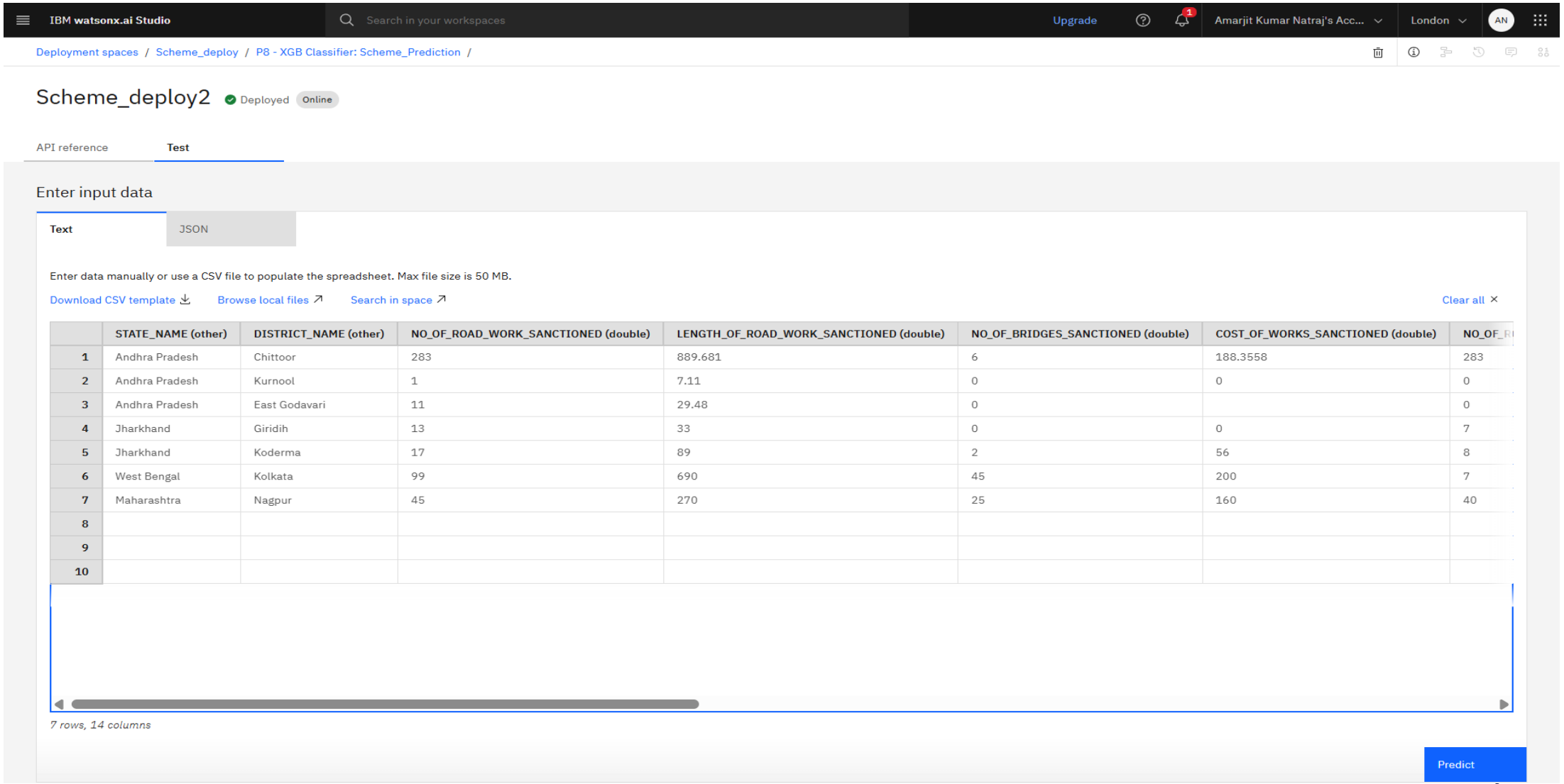

edunet foundation

# RESULT




edunet foundation

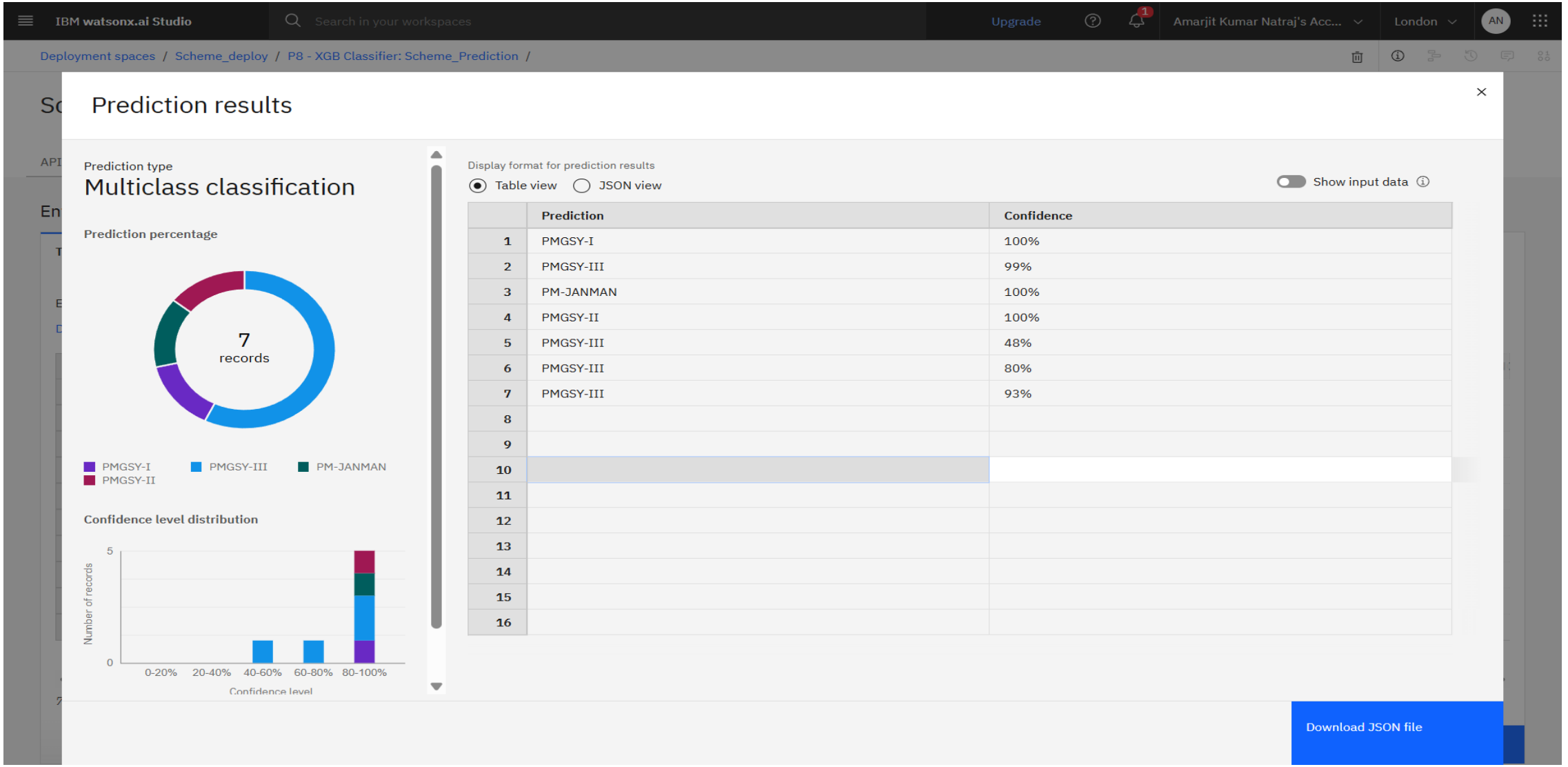
# RESULT




| | STATE_NAME (other) | DISTRICT_NAME (other) | NO_OF_ROAD_WORK_SANCTIONED (double) | LENGTH_OF_ROAD_WORK_SANCTIONED (double) | NO_OF_BRIDGES_SANCTIONED (double) | COST_OF_WORKS_SANCTIONED (double) | NO_OF_R |
|---|---|---|---|---|---|---|---|
| 1 | Andhra Pradesh | Chittoor | 283 | 889.681 | 6 | 188.3558 | 283 |
| 2 | Andhra Pradesh | Kurnool | 1 | 7.11 | 0 | 0 | 0 |
| 3 | Andhra Pradesh | East Godavari | 11 | 29.48 | 0 | | 0 |
| 4 | Jharkhand | Giridih | 13 | 33 | 0 | 0 | 7 |
| 5 | Jharkhand | Koderma | 17 | 89 | 2 | 56 | 8 |
| 6 | West Bengal | Kolkata | 99 | 690 | 45 | 200 | 7 |
| 7 | Maharashtra | Nagpur | 45 | 270 | 25 | 160 | 40 |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |

edunet foundation

# RESULT




| | Prediction | Confidence |
|---|---|---|
| 1 | PMGSY-I | 100% |
| 2 | PMGSY-III | 99% |
| 3 | PM-JANMAN | 100% |
| 4 | PMGSY-II | 100% |
| 5 | PMGSY-III | 48% |
| 6 | PMGSY-III | 80% |
| 7 | PMGSY-III | 93% |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |

edunet foundation

# CONCLUSION

The developed Machine Learning model using **XGBoost Classifier** successfully automates the classification of PMGSY infrastructure projects into their respective schemes with high accuracy. By leveraging **IBM Cloud Lite services** for deployment, the solution ensures scalability, real-time accessibility, and reduces manual effort in project monitoring and classification. This intelligent system aids government bodies and infrastructure planners in making faster, data-driven decisions for transparent budget allocation and efficient resource management.

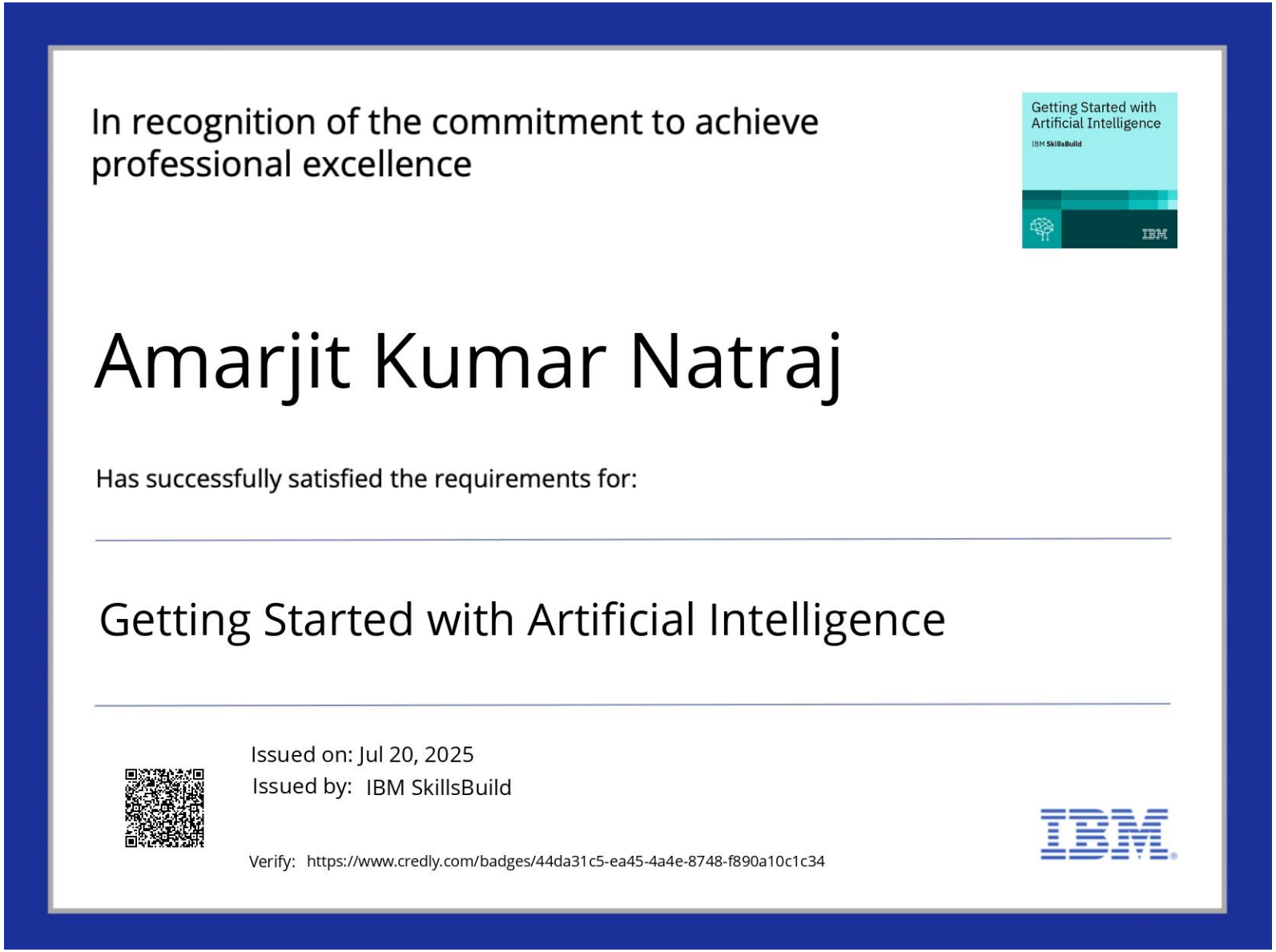
# FUTURE SCOPE

- Incorporate GIS mapping and satellite imagery for better classification.
- Optimize the model with advanced feature selection and tuning.
- Expand the system to cover multiple states and infrastructure schemes.
- Utilize Edge Computing for real-time, on-site predictions.
- Explore Deep Learning techniques for complex data scenarios.
- Implement a feedback loop for continuous model improvement.

# REFERENCES


- **AI Kosh – Pradhan Mantri Gram Sadak Yojana (PMGSY) Dataset**
  Source of project data used for training and evaluating the classification model.
  https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html
- **Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.**
  Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
  DOI: 10.1145/2939672.2939785
- **Scikit-learn: Machine Learning in Python**
  Pedregosa et al., Journal of Machine Learning Research, 2011.
  https://scikit-learn.org/stable/
- **IBM Cloud & Watson Studio Documentation**
  Resources used for deploying the ML model and API on IBM Cloud services.
  https://cloud.ibm.com/docs
- **Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques.**
  Informatica, 31(3), 249-268.
- **Best Practices for Data Preprocessing in Machine Learning**
  Article by Towards Data Science on Medium platform.
  https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825

# IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Amarjit Kumar Natraj

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/44da31c5-ea45-4a4e-8748-f890a10c1c34

IBM

edunet foundation

# IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

Amarjit Kumar Natraj

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 21, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/58c91d30-ddc7-463b-b4de-7f8aea6432f4

IBM

edunet foundation

# IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Amarjit Kumar Natraj

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

edunet foundation

# THANK YOU

edunet foundation