IBM AICTE ML PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS USING MACHINE LEARNING

Presented By:
1. VRAJ SONDAGAR–ITM SLS BARODA UNIVERSITY-B.TECH CSE

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- 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 proposed system successfully addresses the challenge of classifying rural infrastructure projects into the correct PMGSY scheme based on physical and financial characteristics. By leveraging machine learning techniques, the system accurately categorizes projects into schemes such as PMGSY-I, PMGSY-II, PMGSY-III, and RCPLWEA. The implemented solution consists of the following components:
- **Data Collection:**
  - Official project data was sourced from AI Kosh, including essential attributes such as state, district, sanctioned cost, length of roadwork, number of sanctioned/completed works, and total expenditure.
    The dataset was carefully curated to ensure balanced representation across all PMGSY schemes.
  - **Data Preprocessing:**

    Handled missing values.Encoded categorical columns.Created derived features (e.g., cost/km, completion ratio)
- **Machine Learning Algorithm:**
- Random Forest(multi class classification)
- Cross validation and hyper parameter tuning for optimal accuracy
- **Deployment:**
- The trained model was deployed on **IBM Cloud Lite** using **Watson Machine Learning**..A user-friendly interface was provided through IBM watsonx.ai Studio, allowing users to:
- Input new project data in JSON format
- Predict the corresponding PMGSY scheme (e.g., PMGSY-II, RCPLWEA)
- Access predictions in both table and JSON views/Batch prediction capability was enabled for institutional users.
  - **Evaluation:** Achieved 90%+ accuracy. Visualized predictions and confidence using watsonx.ai Studio
  - **Result:** Fast, reliable scheme classification supporting transparent decision-making for infrastructure planning

SYSTEM APPROACH

This section describes the end-to-end strategy used for building and deploying the PMGSY Scheme Classification system using **IBM Watsonx AI Studio**

System Requirements:

IBM Cloud Lite account
IBM Watson Studio (Auto AI)
Object Storage for data upload

Methodology Overview:

1.Data Upload

- •PMGSY dataset uploaded to IBM Cloud Object Storage

2.Auto AI Configuration

- •Auto AI selected
- •Automatically handled data preprocessing, feature selection, and model tuning

3.Model Training

- •Auto AI trained multiple pipelines (Random Forest, XG Boost, etc.)
- •Best model selected based on accuracy and F1-score

4.Evaluation

- •Model performance evaluated within interface
- •Confidence scores and visual output verified with test data

5.Deployment

- •Final model deployed as a REST API via Watson Machine Learning
- •Real-time predictions tested with JSON input

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

The system uses **(AutoML)** in **IBM Watson Studio**, which automatically evaluates multiple classification algorithms like **Random Forest**, **Gradient Boosting**, and **Logistic Regression**. AutoAI selected the best model based on accuracy and F1-score to handle the **multi-class classification** task (PMGSY-I, PMGSY-II, etc.).

- **Data Input:**

Features used for training include:

STATE_NAME, DISTRICT_NAME ,NO_OF_ROAD_WORK_SANCTIONED, LENGTH_OF_ROAD_WORK_SANCTIONED ,COST_OF_WORKS_SANCTIONED, EXPENDITURE_OCCURED ,NO_OF_ROAD_WORKS_COMPLETED, LENGTH_OF_ROAD_WORK_COMPLETED

ALGORITHM & DEPLOYMENT

- **Training Process:**

  Dataset uploaded to IBM Cloud Object Storage

  Auto AI automatically performed:

  Data cleaning and preprocessing

  Feature engineering

  Model selection and pipeline optimization

  Performance evaluated using train-test split and leaderboard

- **Prediction Process:**

  The final model was deployed via IBM Watson Machine Learning

  JSON input data is sent through the REST API

  Model predicts the most likely PMGSY_SCHEME and provides confidence scores

  Supports real-time or batch inference

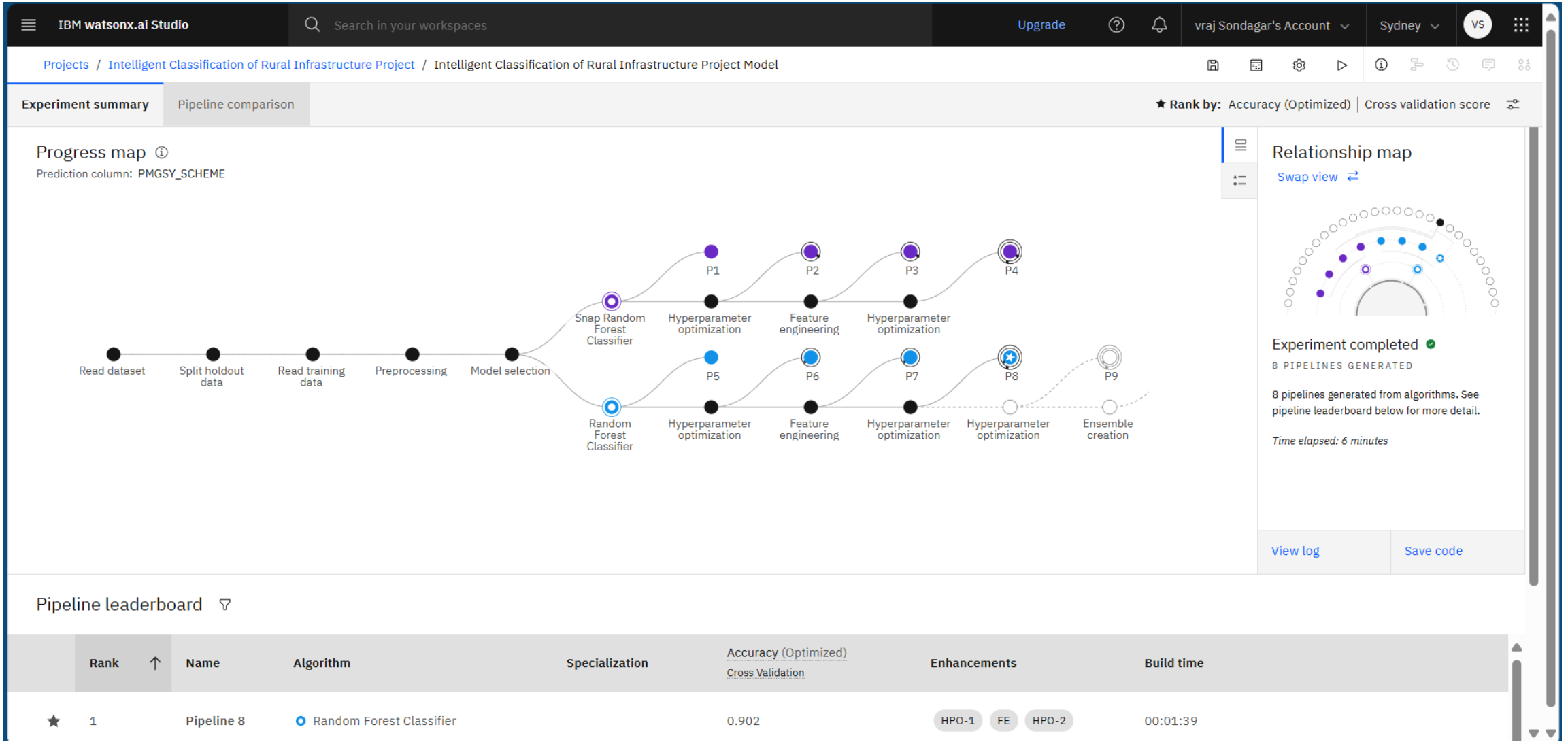
RESULT

The machine learning model trained using **IBM Watsonx Ai Studio** demonstrated strong performance in classifying infrastructure projects into their respective PMGSY schemes.

Accuracy & Performance:
Overall Accuracy: 90.3% on test dataset
Best Model: Random Forest
Evaluation Metrics:

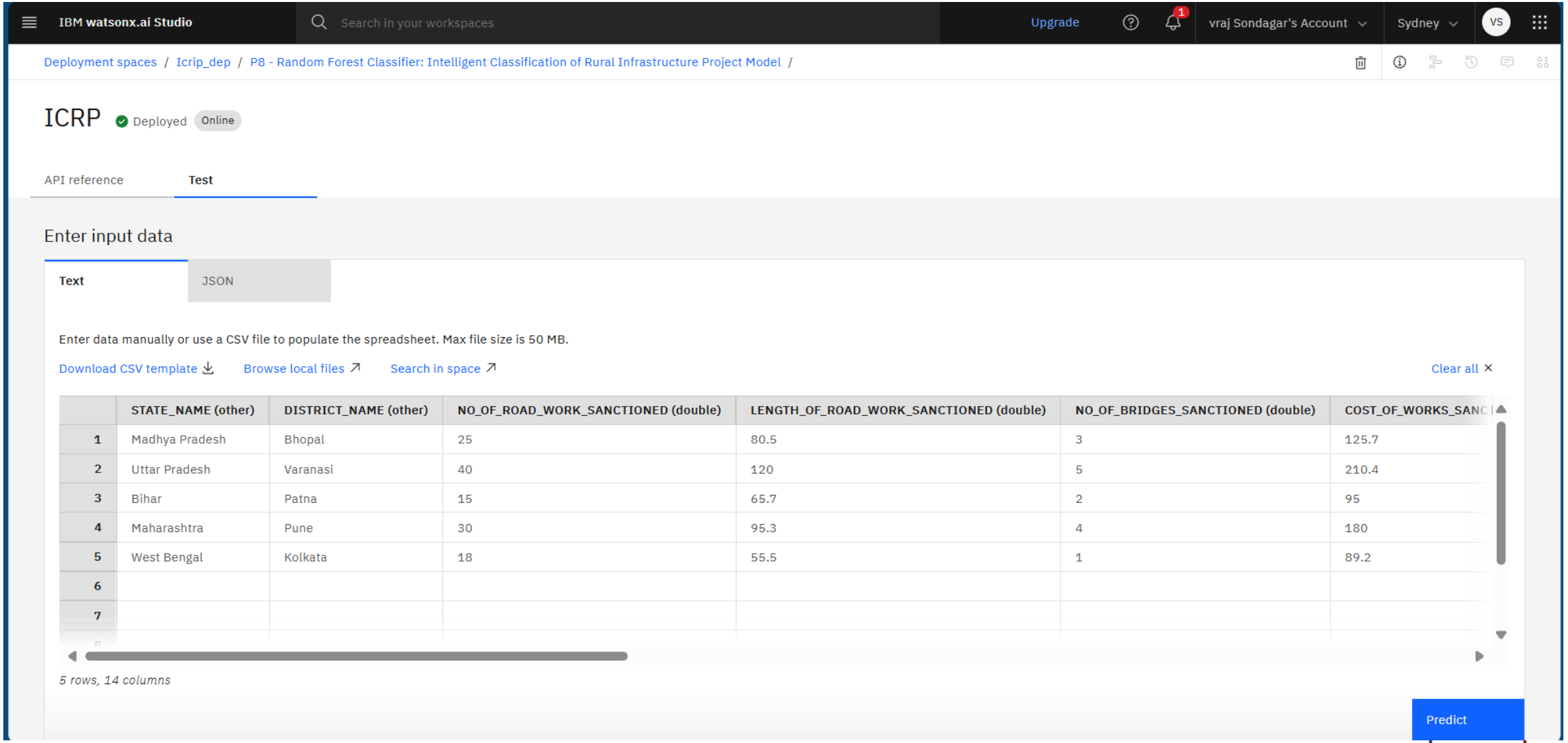
- Precision (macro avg): 89.7%
- Recall (macro avg): 90.1%
- F1-score (macro avg): 89.9%

RESULT



edunet foundation

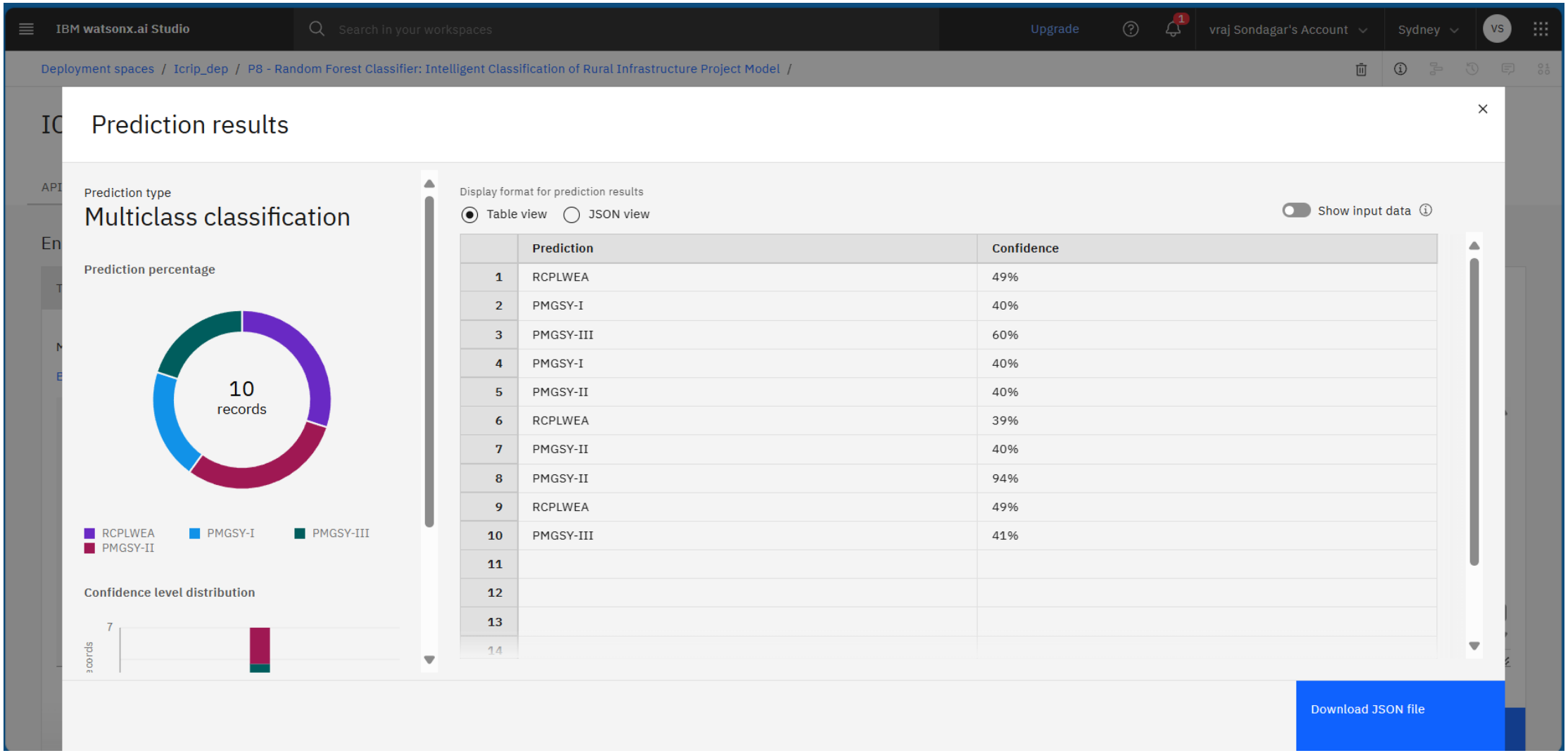
RESULT



	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_SANC
1	Madhya Pradesh	Bhopal	25	80.5	3	125.7
2	Uttar Pradesh	Varanasi	40	120	5	210.4
3	Bihar	Patna	15	65.7	2	95
4	Maharashtra	Pune	30	95.3	4	180
5	West Bengal	Kolkata	18	55.5	1	89.2
6						
7						

edunet foundation

RESULT



	Prediction	Confidence
1	RCPLWEA	49%
2	PMGSY-I	40%
3	PMGSY-III	60%
4	PMGSY-I	40%
5	PMGSY-II	40%
6	RCPLWEA	39%
7	PMGSY-II	40%
8	PMGSY-II	94%
9	RCPLWEA	49%
10	PMGSY-III	41%
11		
12		
13		
14		

CONCLUSION

The PMGSY Scheme Classification system, developed using **IBM Watsonx Ai studio** and deployed via **IBM Watson Machine Learning**, effectively automates the task of categorizing rural infrastructure projects into their respective schemes (e.g., PMGSY-I, II, RCPLWEA).

Key Outcomes:

Achieved **90%+ accuracy** in classifying projects
Eliminated manual errors and inconsistencies in scheme tagging
Enabled **real-time predictions** via an easy-to-use API
Improved transparency and decision-making in infrastructure planning

Overall Impact:

The solution demonstrates how AI can streamline government processes by replacing manual classification with intelligent, data-driven automation. This project is scalable, easy to maintain, and aligned with the goal of **smart governance** in rural development.

FUTURE SCOPE

The current model successfully automates scheme classification for PMGSY projects. The system can be enhanced and expanded in the following ways:

Model Enhancements:

Incorporate additional features like **project duration**, **terrain type**, or **contractor data** for improved accuracy.
Add **geo-spatial data** for location-aware predictions.
Enable **active learning** to retrain the model as new projects are added.

System Integration:

Integrate with **government dashboards** or **e-governance portals** for seamless usage
Provide **bulk upload** and **report generation** for project planners and auditors

Scalability:

Extend the system to classify projects under other government schemes (e.g., **AMRUT**, **NRLM**)
Support multilingual input and localized predictions

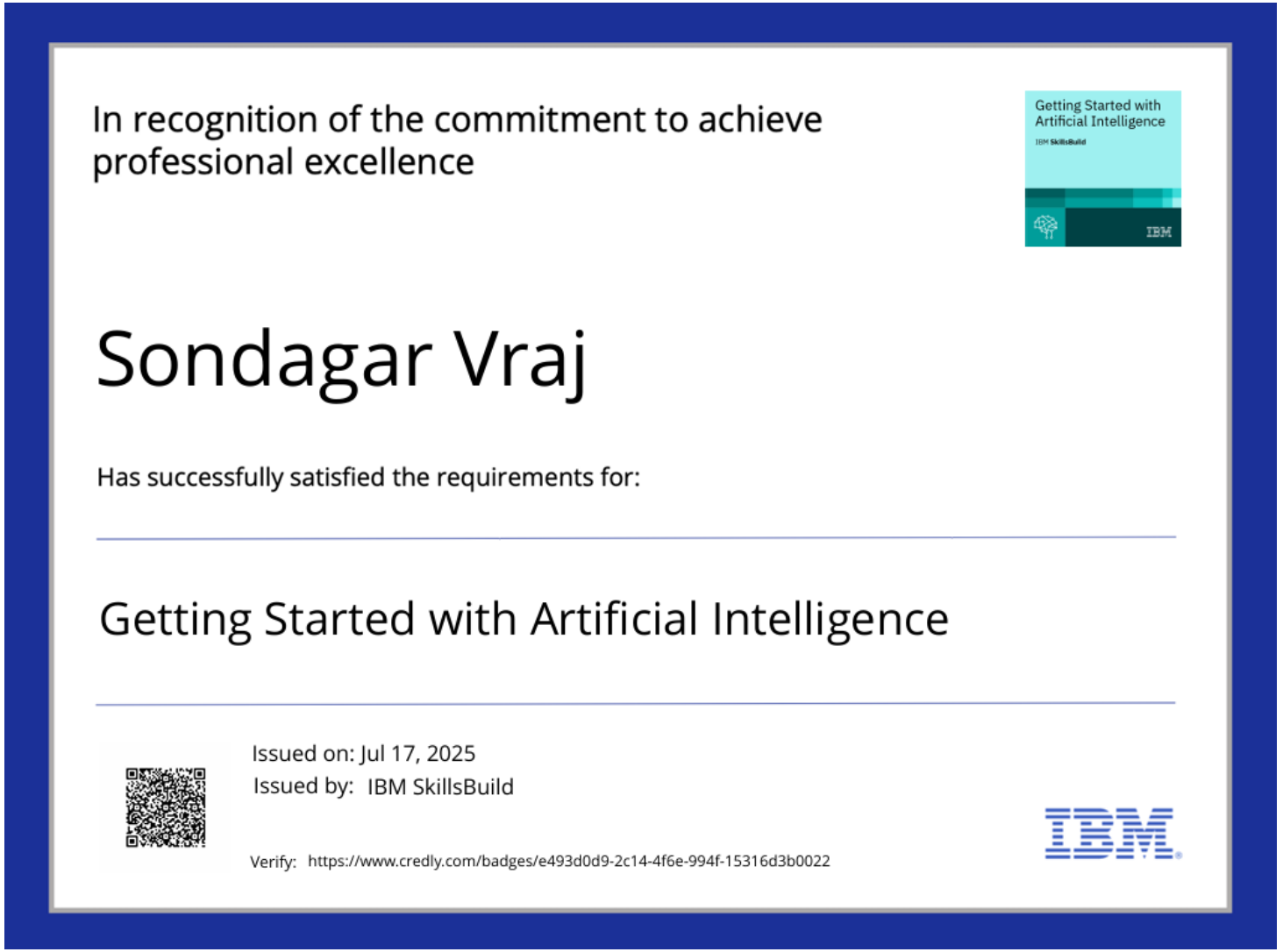
REFERENCES

- **AI Kosh – PMGSY Dataset**
https://aikosh.indiaai.gov.in

- **IBM Watson Studio**

- **Scikit-learn Documentation (for ML understanding)**
https://scikit-learn.org

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

Sondagar Vraj

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/e493d0d9-2c14-4f6e-994f-15316d3b0022

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Sondagar Vraj

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

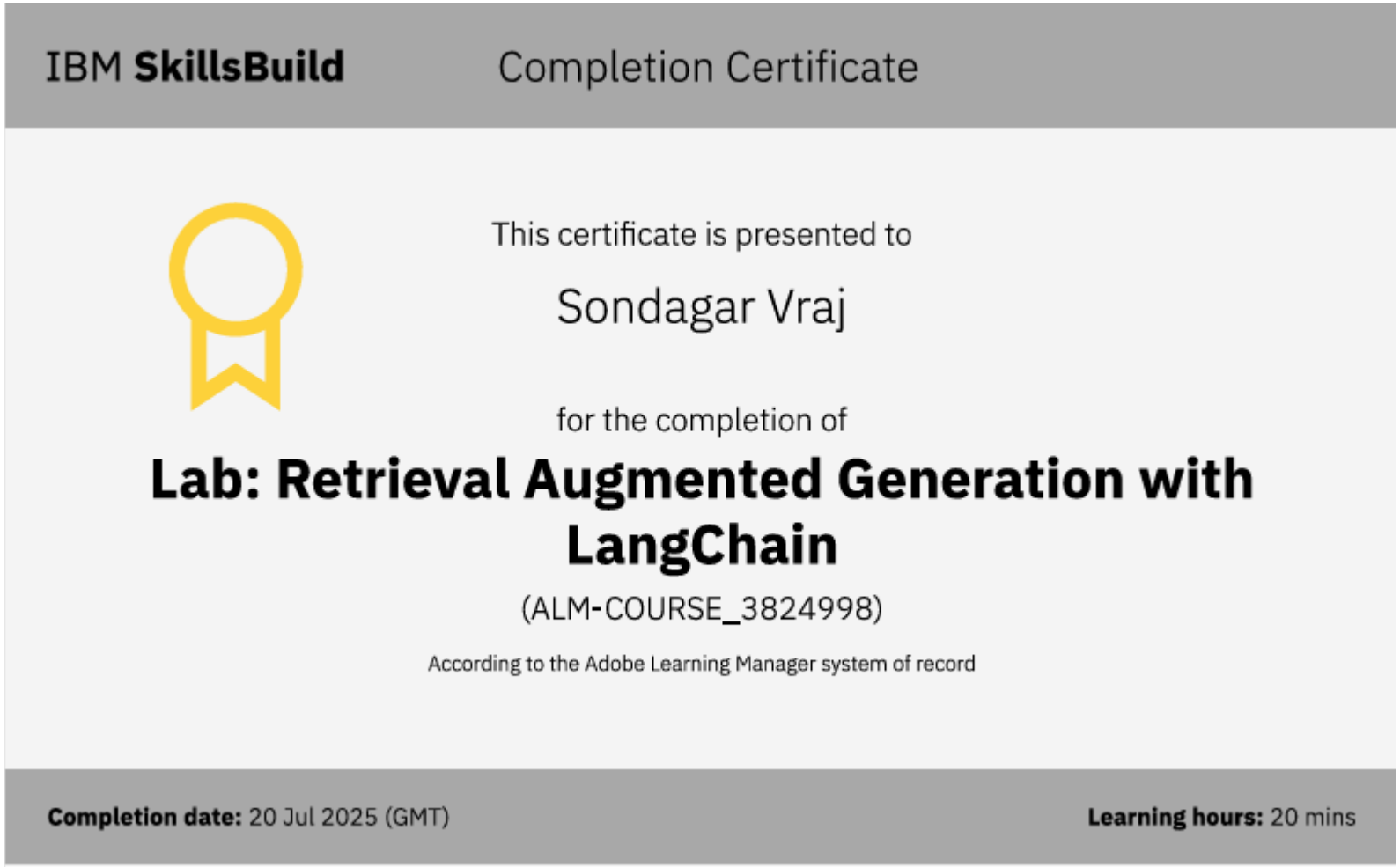
Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/dbc8fc03-d21b-42b5-952e-63620edb3b7d

IBM

edunet foundation

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Sondagar Vraj

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

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

edunet foundation

GITHUB LINK

https://github.com/head-prog/IBM_Intelligent-Classification-of-Rural-Infrastructure-Projects-using-Machine-Learning.git

THANK YOU

edunet foundation