CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By :-
Name - Gyanesh Verma
College - Government Polytechnic Lucknow
Branch - P. G. Diploma in Computer Application

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

Example:

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 aims to address the challenge of predicting that Pradhan Mantri Gram Sadak Yojana (PMGSY) is India's flagship rural development program, providing all-weather road connectivity to unconnected habitations. Efficiently categorizing ongoing and completed projects is crucial for government bodies, infrastructure planners, and policy analysts.
- The solution will consist of the following components:
  - Data Collection:
    - Gather historical data on no. of road work and bridge sanctioned, length of road work and bridge sanctioned, no. of road works and bridge completed, location, and other relevant factors.
  - Data Preprocessing:
    - Preprocess the collected data to handle missing values, outliers, and inconsistencies.
  - Machine Learning Algorithm:
    - Implement a machine learning algorithm, such as a Snap Random Forest Classifier and XGB Classifier to predict PMGSY_SCHEME based on historical patterns.
    - Consider incorporating other factors like Geographical Factors, Technical Specs, Socio-Economic Data and other factors to improve prediction accuracy.
  - Deployment:
    - Develop a user-friendly interface or application that provides real-time predictions for PMGSY_SCHEME at different measures.
  - Evaluation:
    - Assess the model's performance using appropriate metrics such as no. of road work and bridge sanctioned, length of road work and bridge sanctioned, no. of road works and bridge completed, or other relevant metrics.
    - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the PMGSY_SCHEME prediction system based on its physical and financial characteristics.

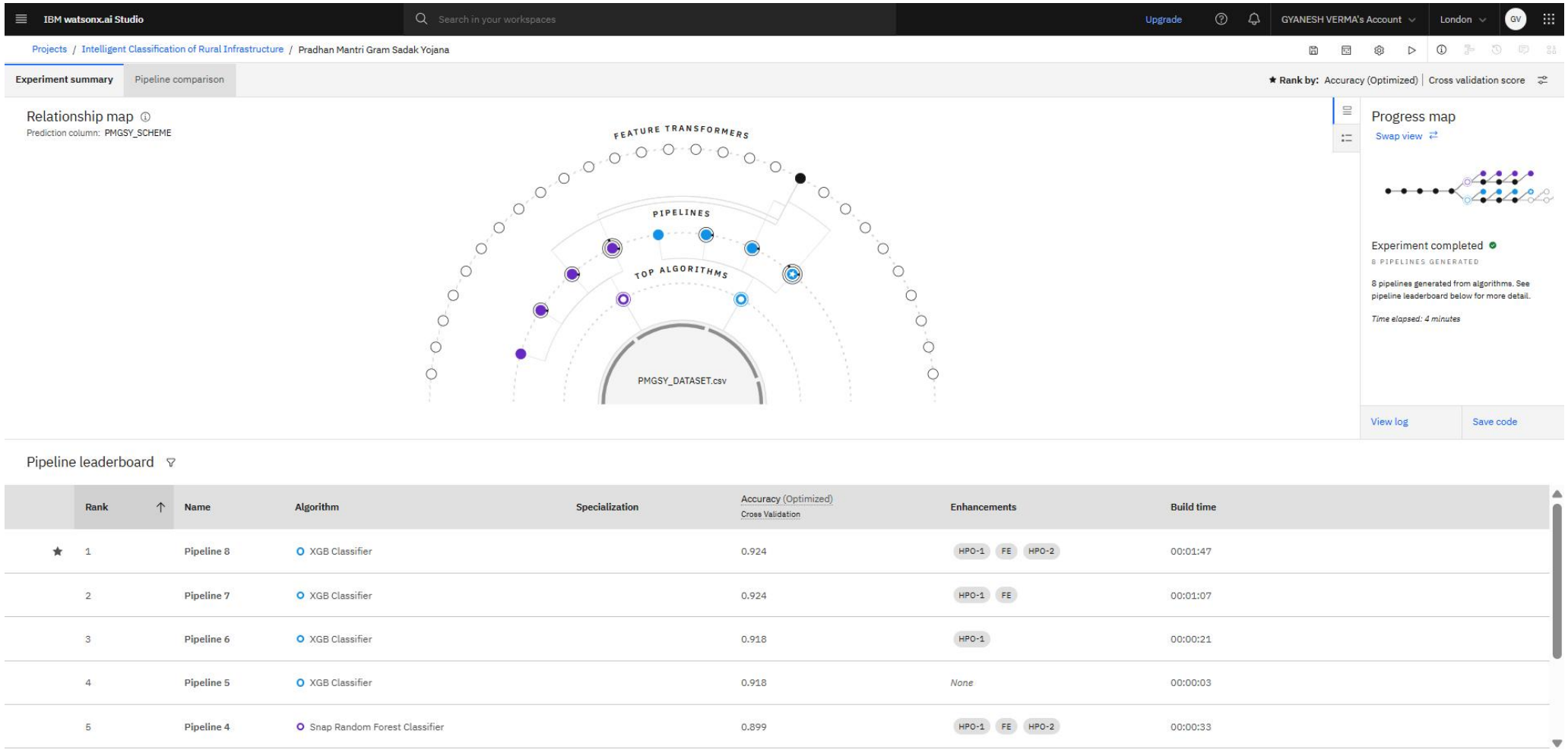
Here's a suggested structure for this section:

- System requirements
  - IBM CLOUD to access cloud services
  - IBM watsonx.ai studio for model development and deployment
  - IBM cloud object storage for data storage

ALGORITHM & DEPLOYMENT

- Algorithm Selection :-
  - **XGB Classifier** - An XGBoost Classifier is a machine learning model used for classification tasks, built upon the Extreme Gradient Boosting (XGBoost) algorithm. XGBoost is an optimized and scalable implementation of gradient boosting, a powerful ensemble learning technique.
- Data Input :-
  - State Name, District Name, No. of Road Work Sanctioned, Length of Road Work Sanctioned, No. of Bridges Sanctioned, Cost of Works Sanctioned, No. of Road Works Completed, Length of Road Work Completed, No. of Bridges Completed, Expenditure Occured, No. of Road Works Balance, Length of Road Work Balance, No. of Bridges Balance.
- Training Process :-
  - The algorithm is trained using historical data, Model selection like Snap Random Forest Classifier and XGB Classifier, Hyperparameter optimization and Feature engineering.
- Prediction Process :-
  - To predict the PMGSY_SCHEME our trained algorithm needs various inputs mentioned in the Data Input section, providing various inputs we can easily predict the scheme involved in Pradhan Mantri Gram Sadak Yojana (PMGSY).

RESULT - RELATIONSHIP MAP

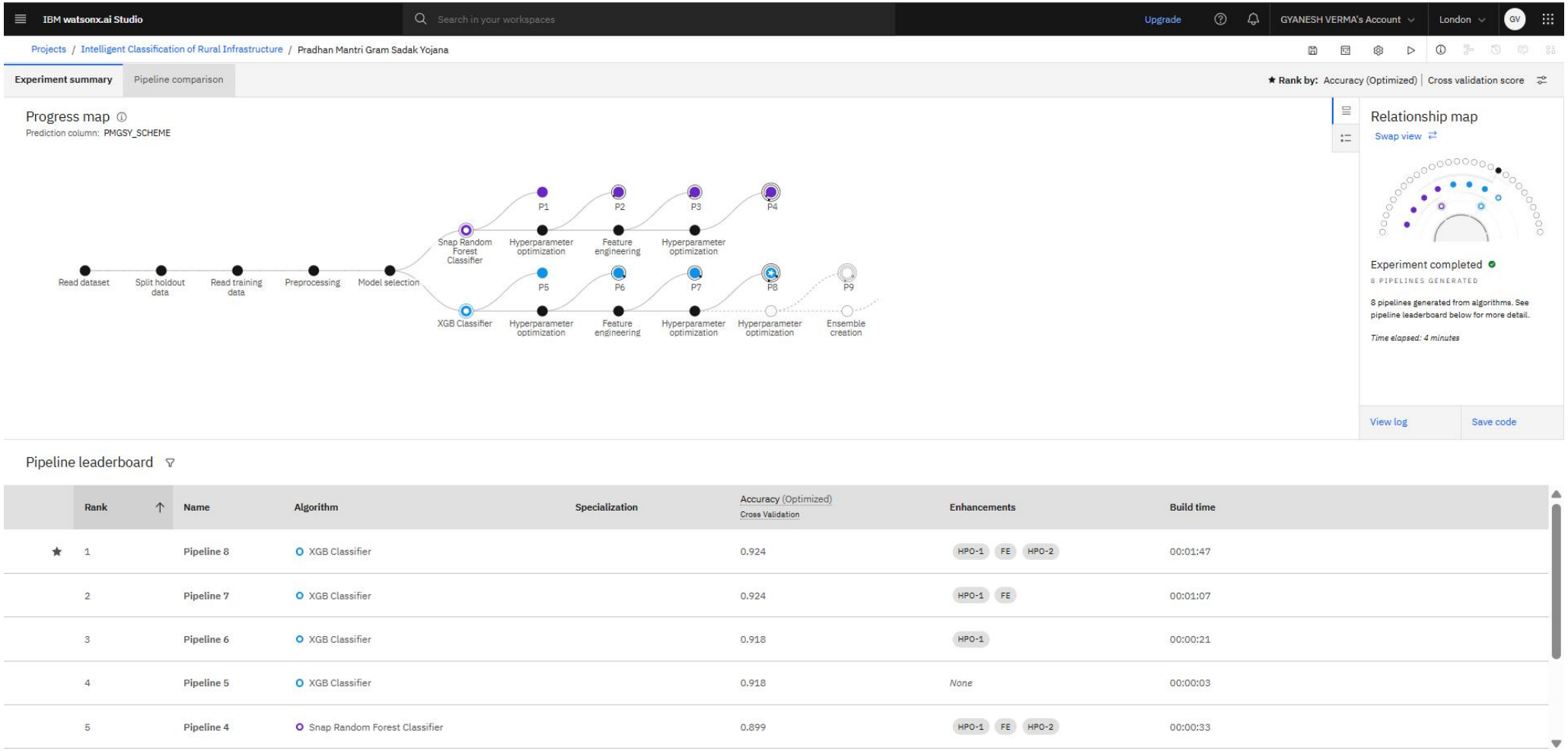


Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 8	XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:47
	2	Pipeline 7	XGB Classifier		0.924	HPO-1 FE	00:01:07
	3	Pipeline 6	XGB Classifier		0.918	HPO-1	00:00:21
	4	Pipeline 5	XGB Classifier		0.918	*None*	00:00:03
	5	Pipeline 4	Snap Random Forest Classifier		0.899	HPO-1 FE HPO-2	00:00:33

edunet foundation

RESULT - PROGRESS MAP

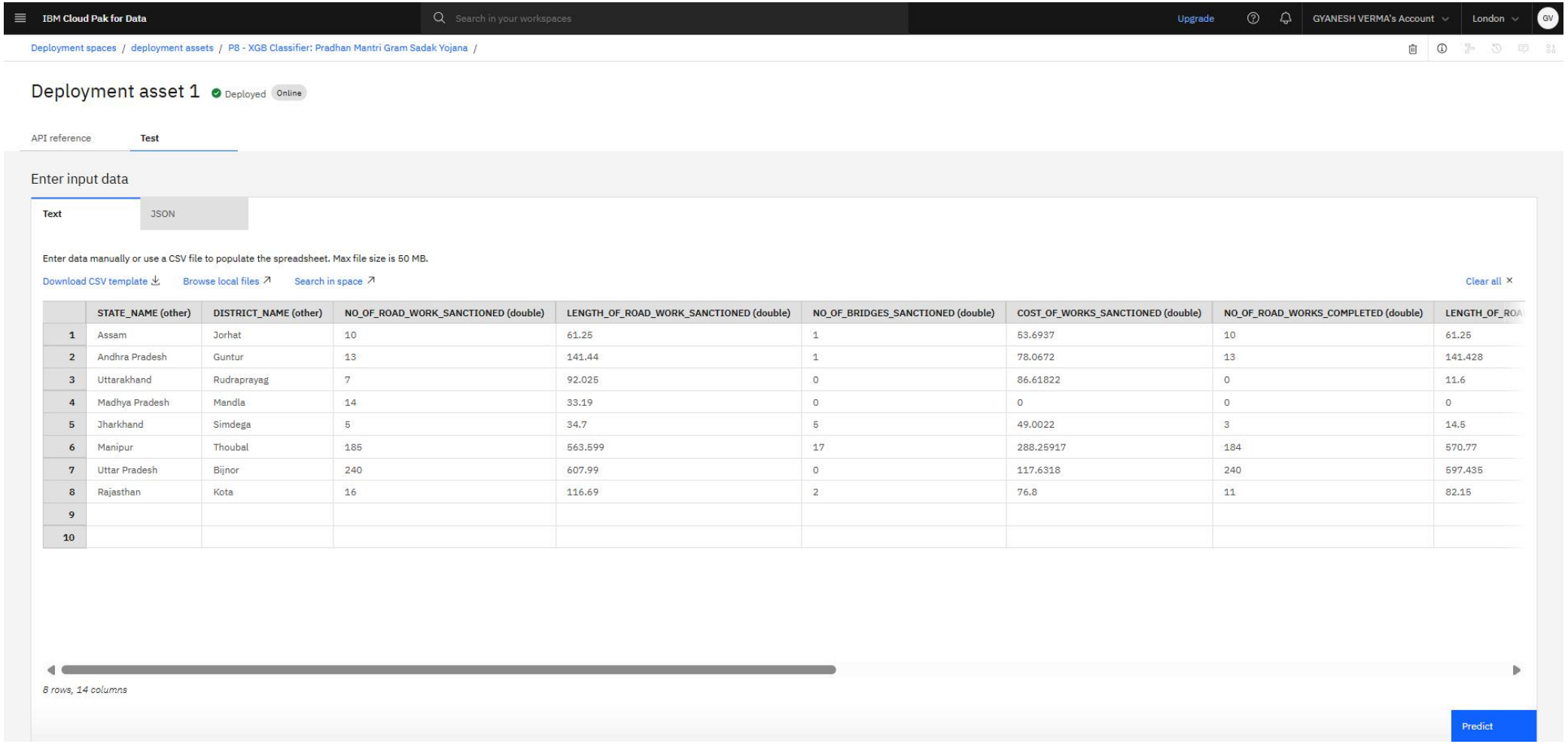


Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 8	XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:47
	2	Pipeline 7	XGB Classifier		0.924	HPO-1 FE	00:01:07
	3	Pipeline 6	XGB Classifier		0.918	HPO-1	00:00:21
	4	Pipeline 5	XGB Classifier		0.918	*None*	00:00:03
	5	Pipeline 4	Snap Random Forest Classifier		0.899	HPO-1 FE HPO-2	00:00:33

edunet foundation

RESULT - DEPLOYMENT ASSETS



	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_ROAD_WORKS_COMPLETED (double)	LENGTH_OF_ROA
1	Assam	Jorhat	10	61.25	1	53.6937	10	61.25
2	Andhra Pradesh	Guntur	13	141.44	1	78.0672	13	141.428
3	Uttarakhand	Rudraprayag	7	92.025	0	86.61822	0	11.6
4	Madhya Pradesh	Mandla	14	33.19	0	0	0	0
5	Jharkhand	Simdega	5	34.7	5	49.0022	3	14.5
6	Manipur	Thoubal	185	563.599	17	288.25917	184	570.77
7	Uttar Pradesh	Bijnor	240	607.99	0	117.6318	240	597.436
8	Rajasthan	Kota	16	116.69	2	76.8	11	82.15
9								
10								

edunet foundation

RESULT - PREDICTION RESULTS

Prediction results

Prediction type

Multiclass classification

Prediction percentage

PMGSY-II PMGSY-III RCPLWEA PMGSY-I

Confidence level distribution

PMGSY-II PMGSY-III RCPLWEA PMGSY-I

Display format for prediction results

Table view JSON view

Show input data

	Prediction	Confidence
1	PMGSY-II	98%
2	PMGSY-II	97%
3	PMGSY-III	100%
4	PMGSY-III	90%
5	RCPLWEA	98%
6	PMGSY-I	100%
7	PMGSY-I	100%
8	PMGSY-III	100%
9		
10		
11		
12		
13		
14		
15		
16		

Download JSON file

edunet foundation

CONCLUSION

- The automatic classification of PMGSY projects into their respective schemes (such as PMGSY-I, PMGSY-II, RCPLWEA, etc.) based on physical and financial characteristics is a vital step toward modernizing infrastructure governance in India. Through this project, we have designed, built, and evaluated a machine learning model capable of accurately categorizing road and bridge construction projects using structured data inputs.
- Key outcomes include:
  - **Increased Efficiency** :- The model significantly reduces manual effort, enabling faster, scalable classification of thousands of projects with high accuracy.
  - **Improved Accuracy** :- Machine learning reduces human error and enhances the reliability of scheme attribution, essential for transparent reporting and budget tracking.
  - **Support for Policy and Planning** :- Automated classification enables real-time analytics for policymakers and infrastructure planners, facilitating evidence-based decisions on resource allocation, scheme performance, and future interventions.
  - **Scalability and Flexibility** :- The model can be continuously updated with new data and extended to incorporate additional schemes or variations in project definitions as policies evolve.

FUTURE SCOPE

- The PMGSY scheme classification system can be enhanced in the following ways:
  - **More Data Sources** : Use satellite, GIS, IoT, and contractor data for better accuracy.
  - **Better Algorithms** : Apply advanced ML models (like XGBoost, deep learning), AutoML, and explainable AI tools.
  - **Wider Reach** : Expand to all Indian states, other schemes, and sectors like housing or electrification.
  - Edge Computing : Deploy models on field devices for offline use; use federated learning for privacy.
  - **System Integration** : Connect with government portals (like OMMAS) via APIs for real-time use.
  - **Smart Learning** : Include human feedback and active learning to keep improving the model.

REFERENCES

Official and Government Sources

- PMGSY Official Website (Ministry of Rural Development)
  - https://pmgsy.nic.in
  - Official guidelines, scheme objectives, data formats, and phase-wise progress reports.
- OMMAS (Online Management, Monitoring and Accounting System)
  - https://omms.nic.in
  - Digital platform for monitoring PMGSY projects, used by field engineers and administrators.
- Bhuvan PMGSY (ISRO GIS platform for PMGSY)
  - https://bhuvan-app3.nrsc.gov.in/bhuvanapp1/bhuvan.jsp?id=pmgsy
  - Satellite/GIS-based project visualization and monitoring tool.
- AI Kosh dataset link –
  - https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

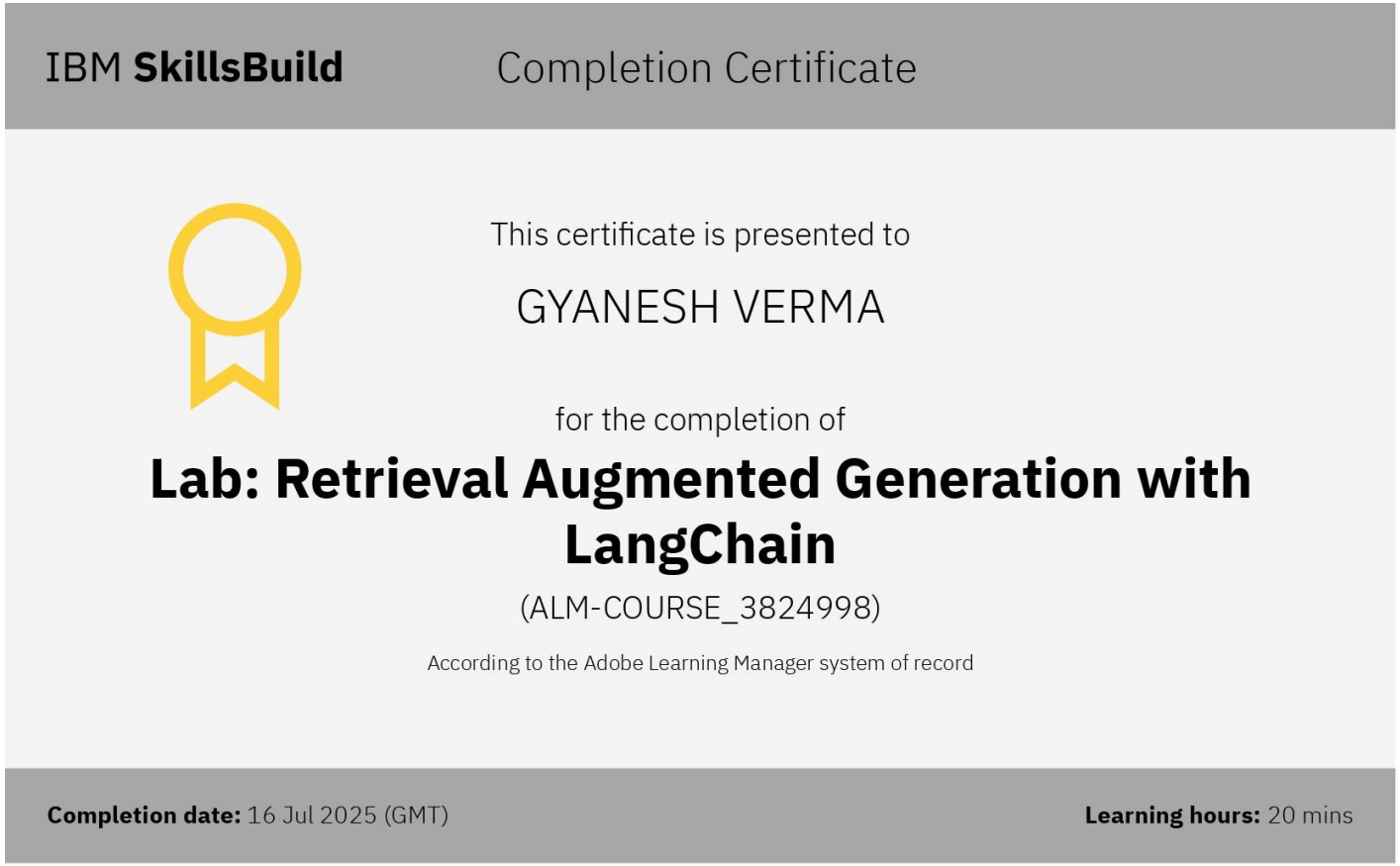
IBM CERTIFICATIONS - GETTING STARTED WITH AI

edunet foundation

IBM CERTIFICATIONS - JOURNEY TO CLOUD

edunet foundation

IBM CERTIFICATIONS - RAG LAB



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

GYANESH VERMA

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 16 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet
foundation