
CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:

1. ADITYA BHADAURIA-Dr. BHIMRAO AMBEDKAR
UNIVERSITY, AGRA-CSE

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 government initiative aimed at providing all-weather road connectivity to rural and remote areas. Over the years, this scheme has been implemented under different phases like PMGSY-I, PMGSY-II, and RCPLWEA — each with its own project specifications, objectives, and funding mechanisms.
- Manually classifying thousands of road and bridge projects into the appropriate PMGSY scheme is a tedious, error-prone, and non-scalable task.
- The goal of this project is to develop a **machine learning model** that can **automatically classify rural infrastructure projects** into their corresponding **PMGSY scheme**, using physical and financial attributes of each project. This will enhance planning accuracy, improve monitoring, and support transparent budget allocation for rural development.

PROPOSED SOLUTION

- Develop a machine learning model that classifies rural infrastructure projects into the correct PMGSY scheme using the provided dataset. The model will analyze financial and physical parameters to determine the scheme category accurately and efficiently, reducing manual errors and improving monitoring.
- **Key Components:**
- **Data Collection:**
Use the AI Kosh dataset containing information on roads and bridges like cost, length, work status, etc.
- **Preprocessing:**
Automatically clean, transform, and engineer features using IBM AutoAI.
- **Model Training:**
Train classification models such as XGBoost using IBM Cloud AutoAI. It selects the best-performing pipeline automatically.
- **Evaluation:**
Validate the model using performance metrics like accuracy, precision, recall, and F1-score.

SYSTEM APPROACH

- The "System Approach" outlines the strategy used to build an ML model that classifies rural infrastructure projects (road/bridge works) under appropriate PMGSY schemes using IBM Cloud services.
- **System Requirements:**
- **IBM Cloud Lite** (mandatory)
- **IBM Watson Studio** – to run AutoAI experiments
- **IBM Cloud Object Storage** – to upload and store the dataset
- **IBM Watson Machine Learning** – to manage and deploy the final model

- **Development Workflow:**

The dataset was uploaded to IBM Cloud Object Storage.

AutoAI in Watson Studio automatically preprocessed the data and trained multiple models.

The goal was to predict the **PMGSY_SCHEME** using project attributes.

AutoAI selected the **XGB Classifier** as the best-performing model.

The trained model was saved and prepared for deployment in Watson Machine Learning.

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

AutoAI in IBM Watson Studio automatically explored multiple ML models.

The **XGB Classifier** was selected based on its superior performance metrics.

- **Input Features:**

The model utilized financial and physical attributes such as road length, cost, bridges, and completion status.

- **Training Process:**

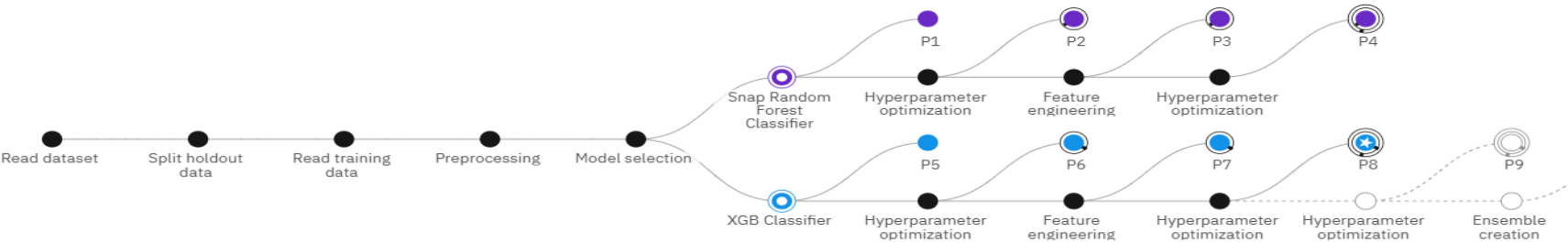
Supervised machine learning was applied to train the model on labeled data under different PMGSY schemes.

- **Deployment:**

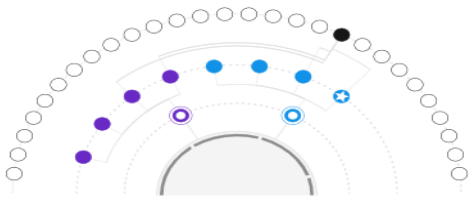
The best model was stored in **Watson Machine Learning** and deployed with an API endpoint to allow real-time project classification.

RESULT

Progress map ⓘ
Prediction column: PMGSY_SCHEME



Relationship map
[Swap view ↔️](#)



Experiment completed ✓
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 6 minutes

[View log](#) [Save code](#)

Pipeline leaderboard ⌵

Rank	↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
------	---	------	-----------	----------------	--	--------------	------------

RESULT

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

ADITYA BHADARIA's Acco...

Dallas

AB

Projects / rural-scheme-predictor / rural-scheme-predictor_ML

Experiment summary

Pipeline comparison

★ Rank by: Accuracy (Optimized) | Cross validation score

Relationship map ⓘ

Prediction column: PMGSY_SCHEME

The diagram illustrates the workflow of the experiment. It starts with a semi-circular base labeled 'PMGSY_DATASET.csv'. From this base, several concentric semi-circular arcs represent different stages of the process. The innermost arc is labeled 'TOP ALGORITHMS' and contains several purple circles. The next arc out is labeled 'PIPELINES' and contains several blue circles. The outermost arc is labeled 'FEATURE TRANSFORMERS' and contains several white circles. Dashed lines connect the circles across the arcs, showing the flow of data and the generation of pipelines from the dataset and algorithms.

Progress map

Swap view ↔

The progress map shows a sequence of steps represented by circles. The first part of the sequence is a straight line of black circles. This is followed by a branching point where the path splits into two parallel paths, each consisting of a series of colored circles (purple, blue, and black). The paths then converge back into a single line of black circles, which ends with a star icon, indicating the completion of the experiment.

Experiment completed ✓

8 PIPELINES GENERATED

8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 6 minutes

View log

Save code

Pipeline leaderboard ⌵

RESULT

Pipeline leaderboard 🔍

	Rank ↑	Name	Algorithm	Specialization	Accuracy (Optimized) <u>Cross Validation</u>	Enhancements	Build time
★	1	Pipeline 8	🔵 XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:49
	2	Pipeline 7	🔵 XGB Classifier		0.924	HPO-1 FE	00:01:11
	3	Pipeline 6	🔵 XGB Classifier		0.918	HPO-1	00:00:22
	4	Pipeline 5	🔵 XGB Classifier		0.918	None	00:00:03

RESULT

rural-scheme-predictor_Deployment

Deployed

Online

API reference

Test

Enter input data

Text

JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

:

Clear all ×

	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF...
1	Andaman And Nicoba	North and Middle Andam	32	60.169	0
2	Andhra Pradesh	Anantapur	619	2169.505	35
3	Bihar	Araria	320	1646.233	220
4	Haryana	Faridabad	12	92.67	0

4 rows, 14 columns

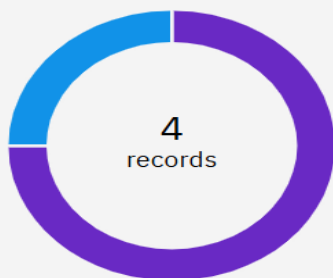
Predict

RESULT

Prediction results

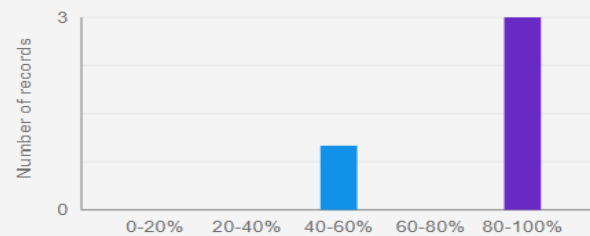
Multiclass classification

Prediction percentage



PMGSY-I PMGSY-III

Confidence level distribution



Display format for prediction results

☒ Table view ☐ JSON view

☐ Show input data ⓘ

	Prediction	Confidence
1	PMGSY-I	96%
2	PMGSY-I	100%
3	PMGSY-I	100%
4	PMGSY-III	52%
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Download JSON file

CONCLUSION

- The project successfully demonstrated the use of IBM Cloud's AutoAI platform to automate the classification of rural infrastructure projects under various PMGSY schemes. By leveraging historical data and machine learning, particularly the XGB Classifier model, we achieved efficient, scalable, and accurate multi-class classification without writing any manual code.
- This solution can significantly aid government agencies and planners by:
 - Enhancing project transparency
 - Streamlining monitoring and analysis
 - Supporting data-driven decision-making for infrastructure development
- The end-to-end deployment on IBM Cloud ensures accessibility, scalability, and real-time prediction capability.

FUTURE SCOPE

- **Model Enhancement:** Future versions can integrate more advanced models or ensemble techniques to improve classification accuracy further. This solution can significantly aid government agencies and planners by:
- **Real-Time Monitoring:** Integrate real-time data feeds to classify new projects as they are added to government systems. Streamlining monitoring and analysis
- **Dashboard Integration:** Build visual dashboards for stakeholders to monitor and analyze scheme-wise performance interactively. The end-to-end deployment on IBM Cloud ensures accessibility, scalability, and real-time prediction capability.
- **Explainability & Transparency:** Use tools like IBM Watson OpenScale to explain model predictions for better trust and compliance.
- **Scalability:** Extend the solution to other rural development schemes beyond PMGSY for broader policy support.

REFERENCES

- **Edunet Foundation** – For organizing the IBM SkillsBuild internship program and providing structured learning resources.
- **IBM SkillsBuild Platform** – For access to IBM Watson Studio, IBM Cloud Object Storage, and AutoAI tools used in model development and deployment.
- **Mentors & Trainers** – Special thanks to the mentors from Edunet and IBM who guided us through IBM AI courses, cloud services, and hands-on projects.
- **AI Kosh Dataset** – Government of India's open data initiative for providing the PMGSY project dataset.
- **Project Guidance & Support** – Thanks to the entire internship support team for continuous motivation and feedback throughout the project.

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Aditya Bhadauria

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/8bbd8915-2e8b-48f5-bf07-26618307c3b0>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Aditya Bhadauria

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



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

Verify: <https://www.credly.com/badges/559a582e-7f86-4bb9-8332-e6770906aed8>



IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to
Aditya Bhadauria

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