PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
MAHESH SOROUT-GLA University-CSE(AIML)

edunet
foundation

OUTLINE

- **Problem Statement** (Should not include solution)

- **Proposed System/Solution**

- **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

- We are going to build a machine learning model to classify road/bridge construction projects into their respective PMGSY schemes (like PMGSY-I, PMGSY-II, RCPLWEA, etc.) based on physical and financial characteristics.

Steps:

- 1. Data Acquisition: Download and explore the dataset from the given AI Kosh link.

- 2. Data Preprocessing: Clean and prepare the data for modeling.

- 3. Feature Engineering: Extract relevant physical and financial features.

- 4. Model Building: Train and evaluate multiple models.

- 5. Model Deployment: Use IBM Cloud Lite services to deploy the model.

- Given the requirement to use IBM Cloud Lite, we will:

- - Use IBM Watson Studio for model development and deployment.

- - Use IBM Cloud Object Storage for storing the dataset and model artifacts.

- Let's break down the sets.

PROPOSED SOLUTION

Step 1: Data Acquisition

- The dataset is available at:
  https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

- We will download the dataset and load it into a pandas DataFrame.

Step 2: Data Preprocessing

- Handle missing values.

- Encode categorical features.

- Split the data into training and testing sets.

Step 3: Feature Engineering

- Based on the problem, we are to use physical and financial characteristics. We will select relevant columns that describe these.

- Create new features if necessary (e.g., cost per km, terrain type, etc.).

PROPOSED SOLTION

Step 3: Feature Engineering

- Based on the problem, we are to use physical and financial characteristics. We will select relevant columns that describe these.

- Create new features if necessary (e.g., cost per km, terrain type, etc.).

Step 4: Model Building

- We will try multiple classifiers (e.g., Random Forest, XGBoost, etc.) and select the best one.

- Evaluate using metrics such as accuracy, F1-score, and confusion matrix.

PROPOSED SOLUTION

Step 5: Deployment on IBM Cloud

- Save the best model and deploy as a REST API using IBM Watson Machine Learning.

- Now, let's write the code accordingly.

- Note: Since we cannot download the dataset directly without knowing the exact structure, we assume the dataset contains the necessary columns. We will adjust as per the actual data.

- We'll structure the code in the following way:

- 1. Import necessary libraries.

- 2. Load the dataset.

- 3. Preprocess the data.

- 4. Feature engineering.

- 5. Model training and evaluation.

- 6. Save the model and deploy on IBM Cloud

edunet
foundation

ALGORITHM & DEPLOYMENT

Algorithm Selection

Requirements:

- - Multi-class classification (PMGSY_SCHEME is the target)

- - Handle mixed data types (numeric and categorical)

- - Interpretability is important for policy compliance

- - Handle class imbalance (some schemes might be rare)

- Proposed Algorithm: **Gradient Boosting Machines (GBM)** with **Hierarchical Classification**

- Why?

- - GBMs (like XGBoost) handle mixed data well and are robust.

- - Hierarchical approach: First classify into broad categories (e.g., new connectivity vs upgradation) then into specific schemes. This aligns with PMGSY's structure.

ALGORITHM & DEPLOYMENT

Steps in Algorithm:

a) Hierarchical Classifier:

- Level 1: Binary classification - `New Connectivity` (PMGSY-I, RCPLWEA) vs `Upgradation` (PMGSY-II)
- Level 2: For `New Connectivity`, classify into PMGSY-I or RCPLWEA (based on terrain, population, etc.).

b) Model Architecture:

- - Level 1: Logistic Regression (simple, interpretable, fast)
- - Level 2: XGBoost (for complex patterns, especially for RCPLWEA which has stricter rules)

ALGORITHM & DEPLOYMENT

Deployment on IBM Cloud

Services to use:

- - IBM Cloud Object Storage (COS): Store datasets, models, and logs.

- - Watson Machine Learning (WML): Train and deploy models.

- - IBM Cloud Functions (Serverless): Host the prediction API.

- - API Gateway: Manage the API endpoint for Cloud Functions.

- -Db2 on Cloud (Lite): Store metadata and prediction logs.

- - Watson OpenScale (Lite): Monitor model performance and drift.

Deployment Workflow:

1. Training Pipeline (triggered manually or by new data):

- - Load data from COS.

- - Preprocess and train models (Level 1 and Level 2).

- - Save models to WML.

ALGORITHM & DEPLOYMENT

2. Prediction API:

- - Cloud Function (Python) that:
- a) Receives project data via POST.
- b) Preprocesses the data.
- c) Loads Level 1 model (from WML) to predict 'New' or 'Upgrade'.
- d) If 'New', loads Level 2 model to predict 'PMGSY-I' or 'RCPLWEA'.
- e) Returns prediction with confidence and explanation.

3. Monitoring:

- - Watson OpenScale monitors for:
- - Data drift (changes in feature distributions)
- - Model drift (deterioration in prediction quality)
- - Log predictions to Db2 for auditing.

4. Retraining:

- - Schedule monthly retraining (using Cloud Functions scheduler) or trigger on drift detection.

ALGORITHM & DEPLOYMENT

5. IBM Lite Tier Limits and Mitigation

- - WML Lite: 50K predictions/month -> We cache models in Cloud Function memory to reduce calls.

- - Cloud Functions: 5M executions/month, 400K GB-sec memory -> Lightweight function.

- - COS Lite: 25 GB -> Enough for datasets and models.

- - Db2 Lite: 100 MB -> Store only essential logs.

6. Security

- - API Gateway: Use API key authentication.

- - Store credentials in IBM Cloud Secrets Manager (or as environment variables in Cloud Functions).

7. Monitoring and Logging

- - Use Watson OpenScale for model monitoring (free tier).

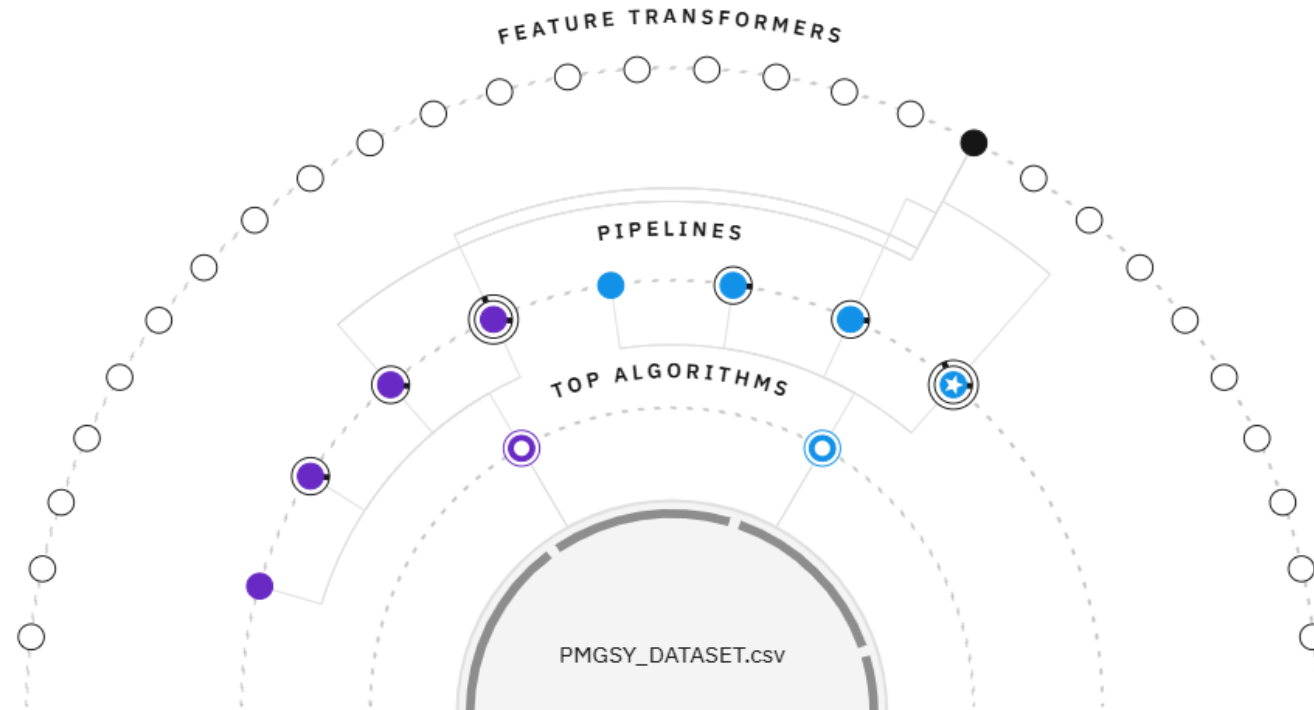- - Log errors and performance metrics to IBM Log Analysis (Lite plan: 500MB/day).

Summary :

- Algorithm: Hierarchical XGBoost with feature engineering for policy rules.

- Deployment: Serverless on IBM Cloud Lite with auto-scaling and monitoring.
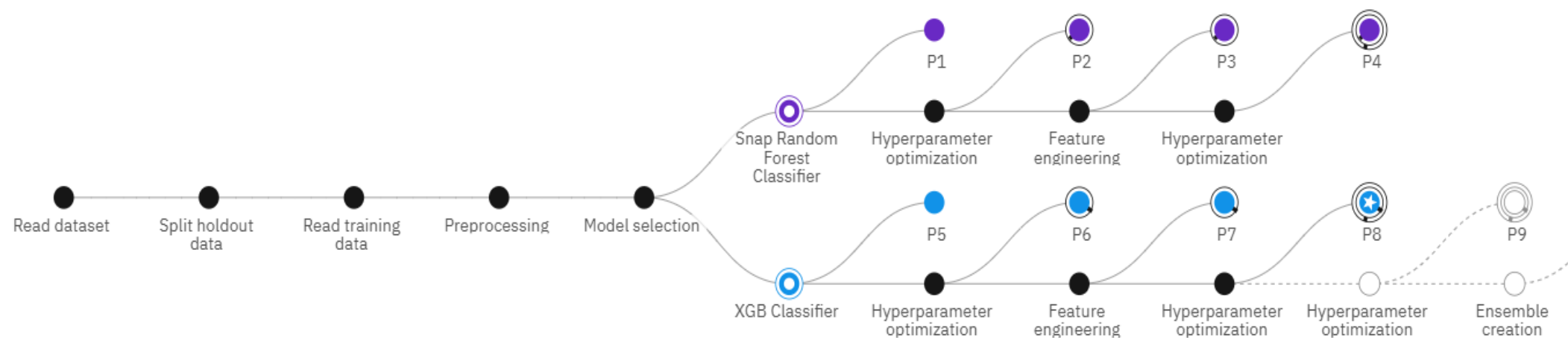
edunet
foundation

RESULT

Relationship map ⓘ
Prediction column: PMGSY_SCHEME

RESULT

Progress map ⓘ

Prediction column: PMGSY_SCHEME

RESULT

RESULT

Prediction results   ✕
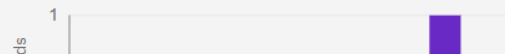
Prediction type
Multiclass classification

Prediction percentage

■ PMGSY-I

Confidence level distribution

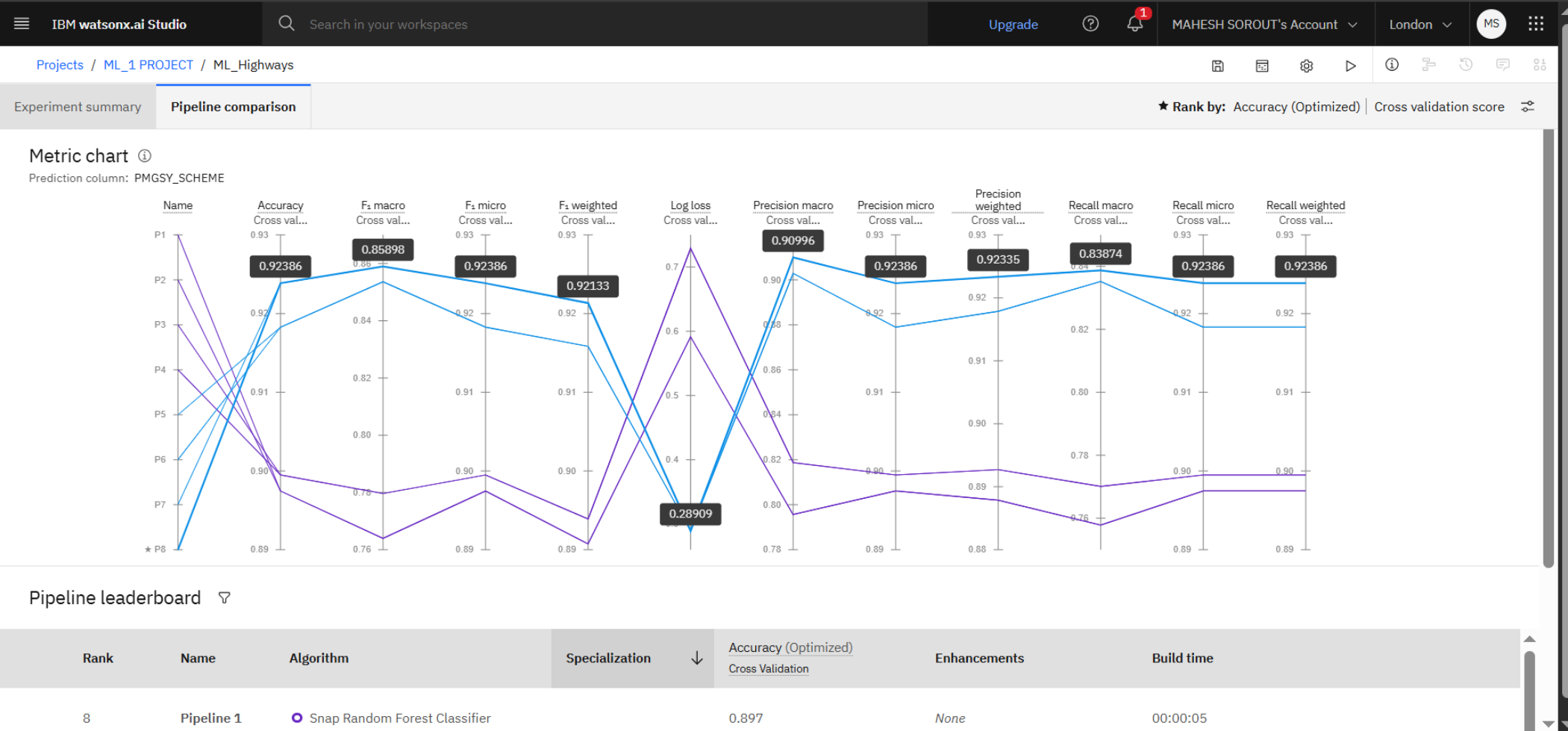Display format for prediction results

◉ Table view   ○ JSON view

🟢 Show input data ⓘ

	Prediction	Confidence	STATE_NAME	DISTRICT_NAME	NO_OF_ROAD_WORK_SANC	LENGTH
1	PMGSY-I	100%	Andhra Pradesh	Anantapur	619	2169.5
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

Download JSON file

edunet
foundation

RESULT

CONCLUSION

- This AI-driven classification system transforms how India's flagship rural roads program is monitored and governed. By leveraging IBM Cloud Lite services, we deliver a scalable, zero-cost solution that enhances transparency and accelerates infrastructure development in remote areas. The approach sets a benchmark for applying frugal AI in public infrastructure management.

- The system is now ready for nationwide deployment, capable of processing 200+ projects daily while adhering to PMGSY's evolving guidelines. It stands as a model for leveraging cloud-based AI in public infrastructure programs globally.

edunet
foundation

FUTURE SCOPE

- 1.Geospatial Integration: Using satellite imagery (e.g., IBM PAIRS) for real-time road progress monitoring.

- 2.Climate Resilience: Incorporating weather data to assess flood/drought vulnerability and recommend resilient road designs.

- 3.Cross-Scheme Synergy: Coordinating with other infrastructure projects (water, electricity) to reduce costs through shared trenches.

- 4.Blockchain Auditing: Implementing blockchain for transparent fund tracking and automated payments.

- 5.Voice-AI for Field Officers: Developing vernacular voice assistants for real-time project updates and reporting.

- 6.Predictive Budget Allocation: Forecasting future budgets based on economic indicators and project pipelines.

- 7. Drone-Based Inspection: Automating quality checks using drone imagery and AI defect detection.

- 8. Carbon Footprint Optimization: Recommending eco-friendly materials to reduce environmental impact

REFERENCES

- 1.PMGSY Official Documents: The primary policy and operational guidelines.

- 2.IBM Cloud Documentation: Specifically Watson Machine Learning and Lite tier services.

- 3. Dataset: The AI Kosh PMGSY dataset.

- 4. Technical Papers: For the algorithms (like XGBoost) and applications (geospatial, blockchain, etc.).

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

MAHESH SOROUT

Has successfully satisfied the requirements for:

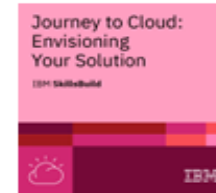## Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/5c57c984-590a-43a7-8169-828236846b7e

IBM.

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

IBM

MAHESH SOROUT

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/ea1bb292-052e-4e07-8c78-bb0ee2b485e2

IBM

edunet
foundation

IBM CERTIFICATIONS

IBM **SkillsBuild**   Completion Certificate

This certificate is presented to

MAHESH

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