CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:

1. Aparna Arunkumar – JAWAHARLALA COLLEGE OF ENGINEERING AND TECHNOLOGY(CSE)

edu**net**
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) aims to provide all-weather road connectivity to rural habitations. Over time, it has evolved into multiple schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.) with distinct objectives. Manually classifying projects into these schemes is inefficient and error-prone. Build and evaluate a machine learning model using IBM Cloud Lite to automatically classify road/bridge projects into the correct PMGSY scheme based on their physical and financial attributes.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of accurately classifying rural infrastructure projects under the correct PMGSY scheme (e.g., PMGSY-I, PMGSY-II, RCPLWEA). This involves leveraging data analytics and machine learning techniques to automate the classification process based on project-specific physical and financial attributes. The solution will consist of the following components:

- Data Collection:

  - Gather historical data on rural infrastructure projects under the PMGSY program, including scheme type, project cost, road and bridge length, completion status, and location-specific details such as state and district.

  - Utilize available data sources such as AI Kosh to obtain consistent records across different PMGSY phases (PMGSY-I, PMGSY-II, RCPLWEA), ensuring diversity in project characteristics for better classification accuracy.

- Data Preprocessing:

  - Clean and pre-process the collected PMGSY data to handle missing values, remove irrelevant columns (e.g., unnamed or empty fields), and resolve any inconsistencies in project records.

  - Perform feature engineering to extract and transform key attributes—such as completion rate, cost per kilometer, and regional indicators—that may significantly influence the classification of a project under the correct PMGSY scheme.

- Machine Learning Algorithm:

  - Implement a supervised machine learning classification model (e.g., Random Forest, SVM, or Gradient Boosting) to categorize projects based on their physical and financial features.

  - **Optimize model performance through feature engineering and hyperparameter tuning, ensuring accurate classification of PMGSY schemes.**

- Deployment:
  - Develop a user-friendly interface or dashboard that displays the predicted PMGSY scheme for each project based on input features.
  - Deploy the solution using IBM Cloud Lite services, ensuring scalability, accessibility, and efficient performance for end-users such as planners and analysts.
- Evaluation:
  - Assess the classification model's performance using metrics such as accuracy, precision, recall, and F1-score.
    - Continuously monitor results and fine-tune the model based on validation feedback to enhance classification accuracy across PMGSY schemes.

  - Result:

  - The model successfully classified rural infrastructure projects into their respective PMGSY schemes with good accuracy. Evaluation metrics such as precision, recall, and F1-score indicated reliable performance. A confusion matrix was used to compare actual and predicted outputs, showing that the model can support decision-making for planners. The results confirm that machine learning can effectively automate project classification, saving time and improving accuracy.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the PMGSY scheme classification system. Here's a suggested structure for this section:

- **System requirements**

- An IBM Cloud account with access to services like IBM Watson Studio, AutoAI, and Watson Machine Learning. A stable internet connection and browser access are sufficient.

- **Tools and Services Used to Build the Model**

  - **IBM Watson Studio** – for project creation, data analysis, and model development

  - **AutoAI** – to automatically preprocess data, select algorithms, and build the best classification model

  - **Watson Machine Learning** – to deploy, manage, and test the trained model

ALGORITHM & DEPLOYMENT

- Algorithm Selection:

  - A classification algorithm like Random Forest is used to categorize rural infrastructure projects into the correct PMGSY scheme based on project details.

- Data Input:

  - Input features include physical and financial details such as project cost, road length, construction type, and state.

- Training Process:

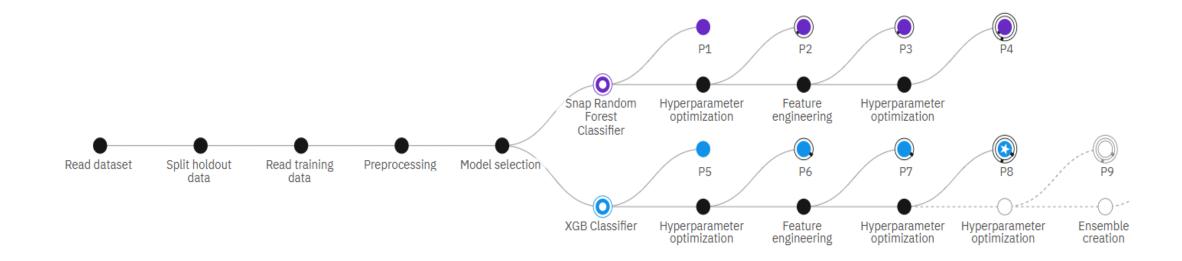 - The model is trained using labeled project data. The data is cleaned, and the model is optimized using techniques like cross-validation.

- Prediction Process:

  - The trained model predicts the PMGSY scheme for new projects based on their input features. Predictions are made through a deployed cloud interface.

RESULT

Progress map ⓘ

Prediction column: PMGSY_SCHEME

RESULT

Relationship map ⓘ
Prediction column: PMGSY_SCHEME

RESULT
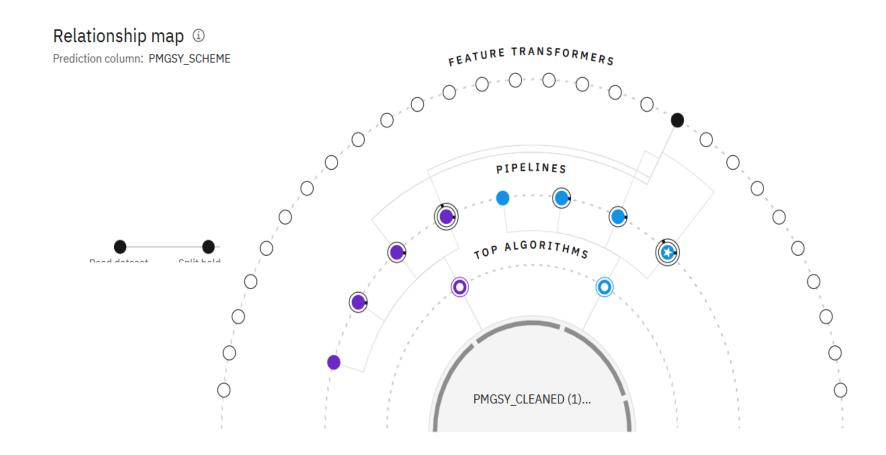
Enter input data

Text | JSON

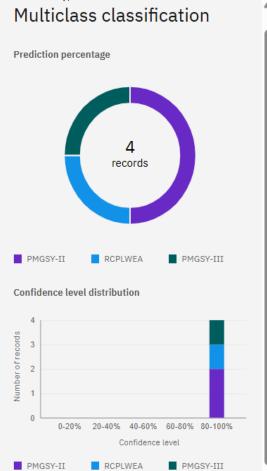Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template ↓   Browse local files ↗   Search in space ↗   Clear all ×

	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_CO
1	Andhra Pradesh	Anantapur	14	125.7	0	56.0711	14
2	Bihar	Aurangabad	29	410.667	12	434.7	23
3	Haryana	Panchkula	1	6.4	0	4.195	1
4	kerala	Thrissur	20	89.8	0	68.83	5
5							
6							
7							
8							
9							

RESULT

RESULT

Prediction results   ✕

Prediction type
Multiclass classification

Display format for prediction results
◉ Table view   ○ JSON view   🔘 Show input data ⓘ

Prediction percentage

	Prediction	Confidence	STATE_NAME	DISTRICT_NAME	NO_OF_ROAD_WORK_SAN(	LENGTH_OF_ROAD_WORK	NO_OF_BRIDGES_SANCTI(	COST_OF_WORKS_SANCTI(	NO_OF_ROAD_W(
1	PMGSY-II	99%	Andhra Pradesh	Anantapur	14	125.7	0	56.0711	14
2	RCPLWEA	90%	Bihar	Aurangabad	29	410.667	12	434.7	23
3	PMGSY-II	96%	Haryana	Panchkula	1	6.4	0	4.195	1
4	PMGSY-III	100%	kerala	Thrissur	20	89.8	0	68.83	5
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									

4 records

■ PMGSY-II ■ RCPLWEA ■ PMGSY-III

Confidence level distribution

■ PMGSY-II ■ RCPLWEA ■ PMGSY-III

CONCLUSION

- The proposed solution effectively classifies rural infrastructure projects into their respective PMGSY schemes using machine learning techniques. The model demonstrated good accuracy and reliability, making it a useful tool for planners and policymakers. Challenges faced during implementation included handling missing or inconsistent data and selecting the most relevant features. Future improvements may include using larger datasets and integrating advanced models for better performance. Accurate classification is essential for effective monitoring, budget allocation, and evaluating the long-term impact of government infrastructure programs.

FUTURE SCOPE

- The system can be improved by adding more data sources like satellite images and local demographics. Better algorithms can be used to increase accuracy. It can be expanded to cover more regions and types of projects. New technologies like edge computing and deep learning can help make the system faster and smarter.

REFERENCES

- Pradhan Mantri Gram Sadak Yojana (PMGSY) Dataset. AI Kosh. Available at: https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

- IBM Cloud. *Watson Studio and Watson Machine Learning*. IBM Cloud Lite . Available at: https://cloud.ibm.com

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with
Artificial Intelligence

IBM SkillsBuild

IBM

APARNA ARUNKUMAR

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/2f18c98f-40b1-4a71-8572-83af7e384a4e

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

APARNA ARUNKUMAR

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

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

IBM

Verify: https://www.credly.com/badges/81ba47ca-41e3-484c-bf96-0c2bf0890682

edunet
foundation

IBM CERTIFICATIONS

THANK YOU