CAPSTONE PROJECT

Intelligent Classification of Rural Infrastructure Projects

Presented By:
Renuka Wadetwar
Symbiosis Institute of Technology, Nagpur
Computer Science

edunet foundation

OUTLINE

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

Problem Statement

The **Pradhan Mantri Gram Sadak Yojana (PMGSY)** is an Indian government initiative focused on rural road infrastructure development. The projects under PMGSY are distributed across multiple schemes such as **PMGSY-I, PMGSY-II, PMGSY-III, RCPLWEA, and PM-JANMAN**, each with varying characteristics. With thousands of records of rural projects, manual classification of these into appropriate schemes is:

- Time-consuming
- Error-prone
- Not scalable

Proposed Solution

The proposed system aims to automate the classification of rural road and bridge construction projects under the PMGSY (Pradhan Mantri Gram Sadak Yojana) schemes. This classification is essential for tracking project implementation, budget allocation, and policy analysis. The system includes the following components:

Data Collection:

- Used the AI Kosh dataset containing historical information about PMGSY projects.
- Dataset features include State, District, Scheme Type, Cost, Sanctioned Work, Completed Work, Expenditure, etc. and stored the data securely on IBM Cloud Object Storage (COS).

Data Preprocessing:

- Handled missing values and dropped irrelevant columns.
- Performed label encoding categorical features and cleaned and structured the dataset for model training.

Machine Learning Algorithm:

- Implemented a Random Forest Classifier to classify schemes.
- Model predicts among: PMGSY-I, PMGSY-II, PMGSY-III, RCPLWEA, PM-JANMAN.

Deployment:

- The trained Random Forest model was saved and uploaded to IBM Watsonx.ai.
- It was deployed in a Deployment Space and exposed via REST API for predictions.

Evaluation:

- Evaluated using accuracy, F1-score, and confusion matrix.
- Model achieved ~89% accuracy with strong precision and recall on major classes.
- Confusion matrix and heatmaps confirmed effectiveness, with scope for improving minority class performance.

System Approach

System Requirements:

- IBM Cloud Lite (Watsonx.ai Studio, Cloud Object Storage, Deployment Space)
- Web browser (IBM Cloud dashboard access)
- Python 3.x runtime environment (Watsonx default)
-

Required Libraries:

- pandas, numpy – for data manipulation and preprocessing
- seaborn, matplotlib – for data visualization
- scikit-learn – for model building and evaluation
- joblib – to save/load trained model
- ibm_boto3, ibm_botocore – to connect with IBM Cloud Object Storage

Algorithm & Deployment

Algorithm Selection:

- Random Forest Classifier was chosen for its robustness and ability to handle tabular data with both categorical and numerical features.

Data Input:

- Model input features include road length, cost, sanctioned and completed work counts; target is the PMGSY scheme type.

Training Process:

- Data was cleaned, encoded, split into train-test sets, and trained using 100 estimators with fixed random state for consistency.

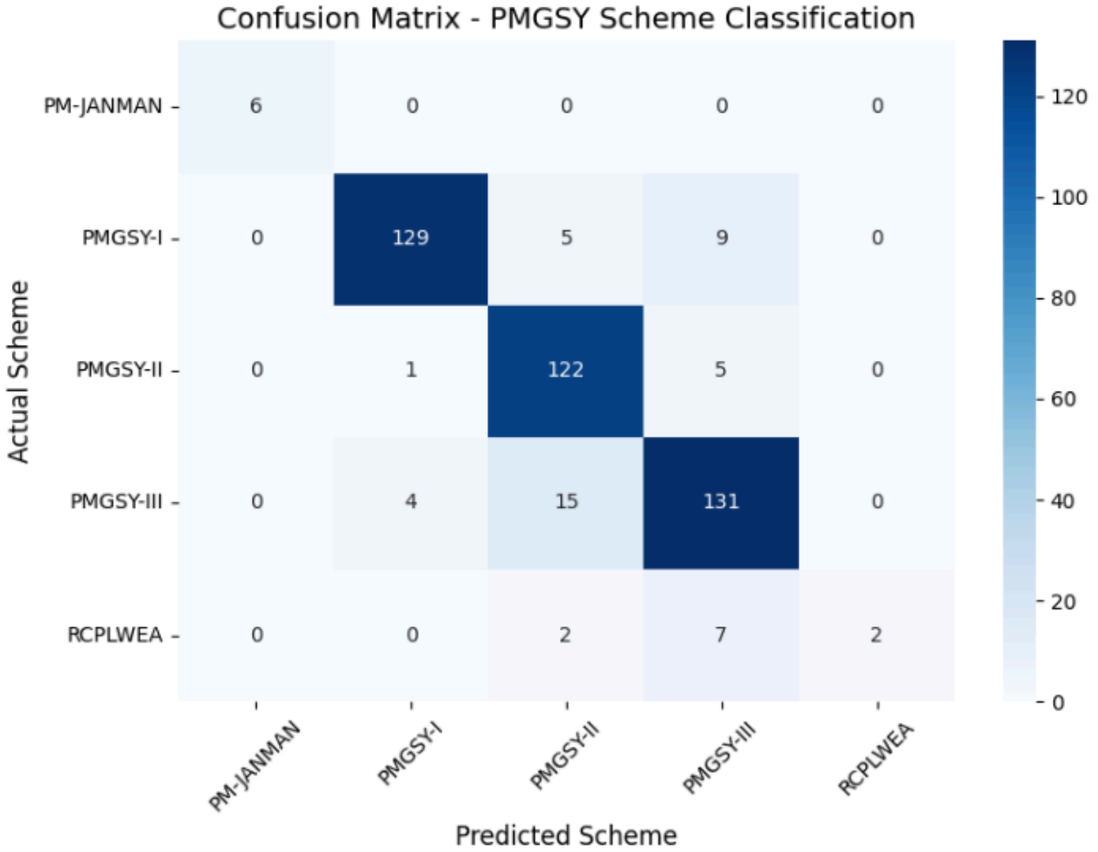
Prediction Process:

- Model predicts the scheme for new data based on trained patterns.

Result

The Random Forest model achieved ~89% accuracy, with strong precision and recall across major PMGSY schemes.

Confusion matrix visualization using Seaborn heatmap clearly showed classification performance, though prediction for minor classes like RCPLWEA was weaker.




```
[18] print("Accuracy Score:", accuracy_score(y_test, y_pred))
```

```
Accuracy Score: 0.8904109589041096
```

edunet foundation

Conclusion

- The classifier successfully automated the categorization of rural infrastructure projects with ~89% accuracy, aiding faster scheme identification.
- Challenges included handling class imbalance (e.g., RCPLWEA), but the model proved effective for large-scale, policy-oriented deployment.

Future scope

- The system can be enhanced by integrating geospatial and satellite data to improve classification accuracy across districts.
- Advanced techniques like ensemble learning or edge deployment can enable real-time, scalable rural infrastructure monitoring.

IBM Certifications

edunet foundation

IBM Certifications

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

Renuka Wadetwar

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/acf4b1ec-be6e-43c0-9036-828fa168615b

IBM

edunet foundation

IBM Certifications

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Renuka Wadetwar Renuka

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 17 Jul 2025 (GMT)

Learning hours: 20 mins

edunet foundation

THANK YOU

edunet
foundation