IBM AICTE PROJECT

PMGSY_35

Presented By:
Sakdasariya Krisha
Sarvajanik Collage of engineering &Technology
Computer Engineering

edunet
foundation

OUTLINE

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

PROBLEM STATEMENT-35

Intelligent Classification of Rural Infrastructure Projects:

Your challenge is 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 accurately classifying rural infrastructure projects into their correct PMGSY schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.) to enable efficient fund allocation and performance monitoring. This involves leveraging machine learning and government-approved data standards to automate classification decisions based on project characteristics. The solution will consist of the following components:

- **Data Collection:**
  - Gather historical PMGSY project data including state, district, road/bridge length, estimated cost, funding source, and terrain type.
  - Integrate external data sources where relevant (e.g., weather patterns, terrain classification, geographic accessibility indices) to enhance classification accuracy
  - **Data Preprocessing:**
  - Clean and preprocess collected data to handle missing values, outliers, and inconsistencies.
  - Apply feature engineering to derive additional relevant attributes such as **cost per kilometer**, **terrain difficulty scores**, or **funding ratios** that may influence classification.
  - **Machine Learning Algorithm:**
  - Implement supervised machine learning models such as **Random Forest, Logistic Regression, or Gradient Boosting** for multi-class classification.
  - Incorporate explainable AI techniques (e.g., feature importance analysis) to help policymakers understand the basis of classification decisions
  - **Deployment:**
  - Develop a user-friendly interface within IBM Watson Studio for uploading new project details and receiving real-time classification predictions.
  - Store and manage datasets securely using **IBM Cloud Object Storage** to support scalability and compliance.
  - **Evaluation:**
  - Assess the model’s performance using metrics such as **Accuracy, Precision, Recall, and F1-Score**.
  - Validate the model on recent unseen project data to ensure reliability in real-world scenarios.
  - **Result:**The proposed solution will provide **accurate, scalable, and automated classification of PMGSY projects**, enabling policymakers to ensure transparent fund allocation, efficient monitoring.

edunet foundation

SYSTEM APPROACH

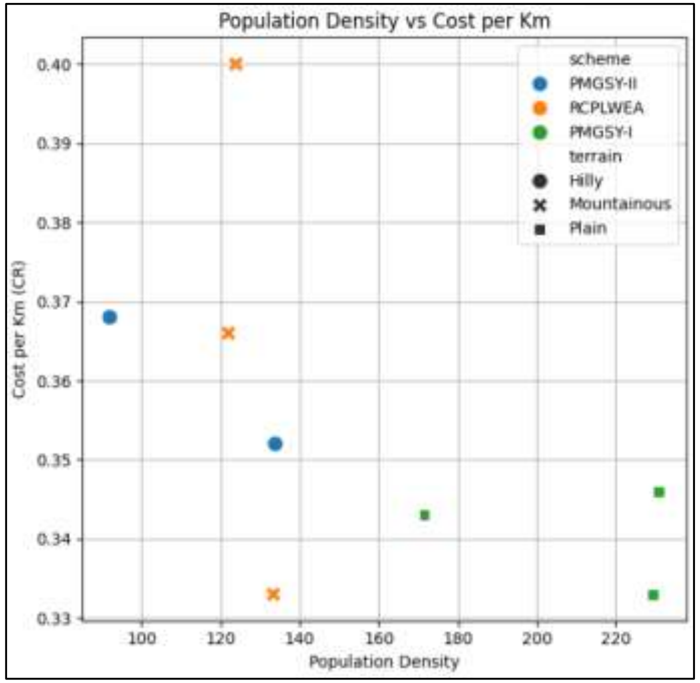
The **System Approach** section outlines the overall strategy and methodology for developing and implementing the **Intelligent Classification of Rural Infrastructure Projects** system. This system leverages machine learning to automatically classify road and bridge projects into their respective PMGSY schemes, enabling efficient fund allocation and performance monitoring.

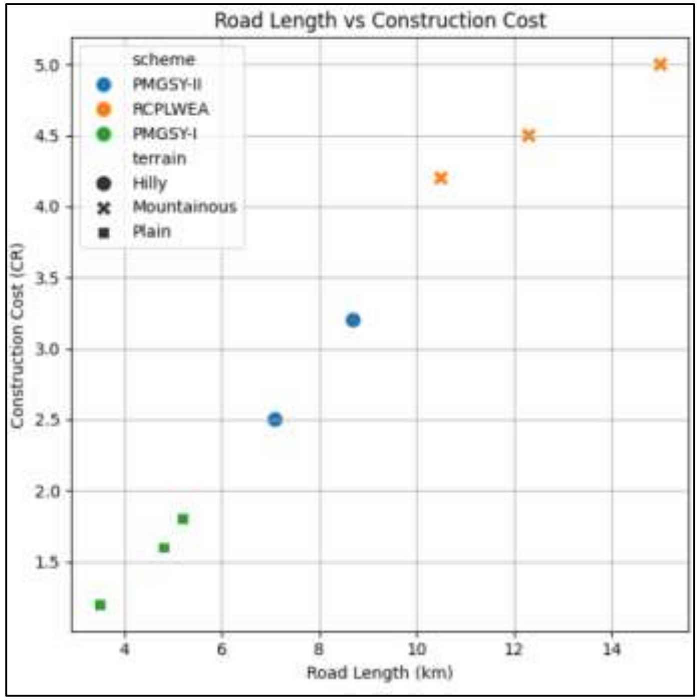
- System requirements
- Hardware Requirements:
- Processor→ Intel i5 or above, Internet connectivity for accessing IBM Cloud services
- RAM:→8 GB minimum (16 GB recommended), Storage→ 20 GB free space
- Software Requirements:
- IBM Cloud account (Lite plan sufficient), IBM Watson Studio, IBM Watson Machine Learning, IBM Cloud Object Storage.
- Library required to build the model
- Data Handling & Preprocessing
- Pandas → For dataset loading and manipulation , numpy → For numerical operations , scikit-learn → For preprocessing , model training, and evaluation
- Visualization
- Matplotlib → For plotting graphs , seaborn → For heatmaps and advanced visualizations.
- Machine Learning Models
- Scikit-learn → Logistic Regression, Random Forest, Gradient Boosting , Joblib → For saving and loading trained models
- Deployment
- Ibm-watston-machine-learning →To deploy the trained model on IBM Watson Machine Learning , request →For making REST API calls (optional for integration)

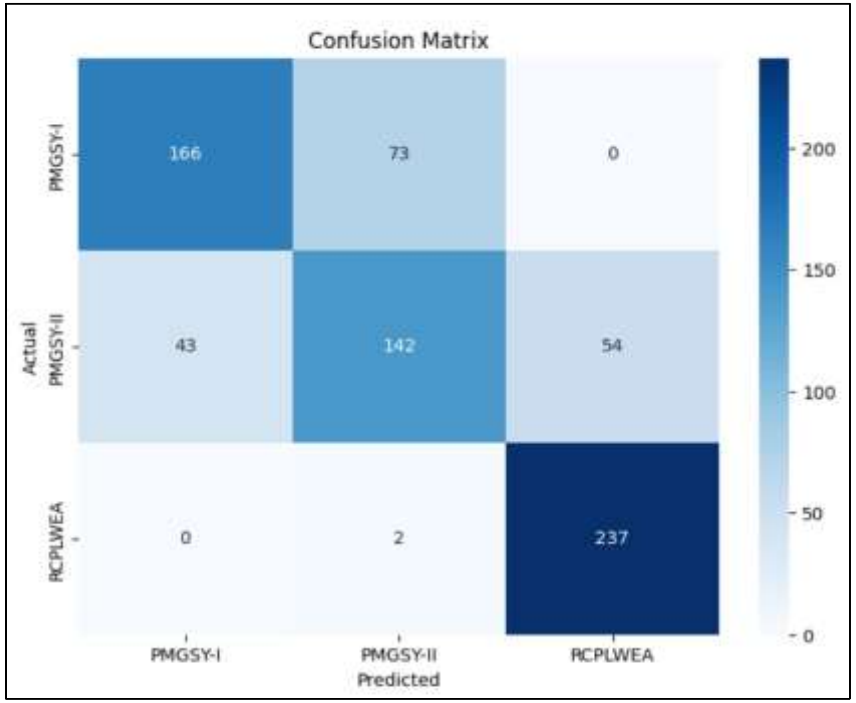
ALGORITHM & DEPLOYMENT

- Algorithm Selection:
  - For this project, the **XGBoost Classifier** was selected due to its high accuracy, efficiency, and ability to handle structured tabular data.
  - Reason for Selection:
  - Supports multi-class classification (required for PMGSY-I, PMGSY-II, RCPLWEA). Handles both categorical (terrain type) and numerical (cost, road length, population) features. Provides feature importance scores, aiding transparency in government decision-making.
- Data Input:
  - **Road length (km)** → Length of the road constructed , **Construction cost (in Crores)** → Total project cost , **Terrain type** → One-hot encoded (Plain, Hilly, Mountainous) , **Population served** → Number of beneficiaries , **Cost per Km** (cost / road length) , **Population Density** (population served / road length)
- Training Process:
  - Data Preprocessing: One-hot encoding for categorical variables (terrain) , Scaling of numerical features using StandardScaler. , Handling class imbalance with **SMOTE (Synthetic Minority Oversampling Technique).**
  - Training: Train-test split: **80% training, 20% testing.**
  - Evaluation Metrics: **Accuracy Score** (for overall correctness) , **Classification Report** (Precision, Recall, F1-Score) , **Confusion Matrix** (visualization for scheme-wise performance.
- Prediction Process:
  - The trained model takes user/project inputs (road length, cost, terrain, population) , Features are processed (cost/km & population density calculated, scaled, and encoded).
  - The model outputs: **Predicted PMGSY Scheme** (PMGSY-I, PMGSY-II, RCPLWEA) , **Confidence Scores** (probabilities for each scheme).
  - Visualization: **Top 5 Feature Importance Graph** to show which factors most influenced the prediction.

RESULT





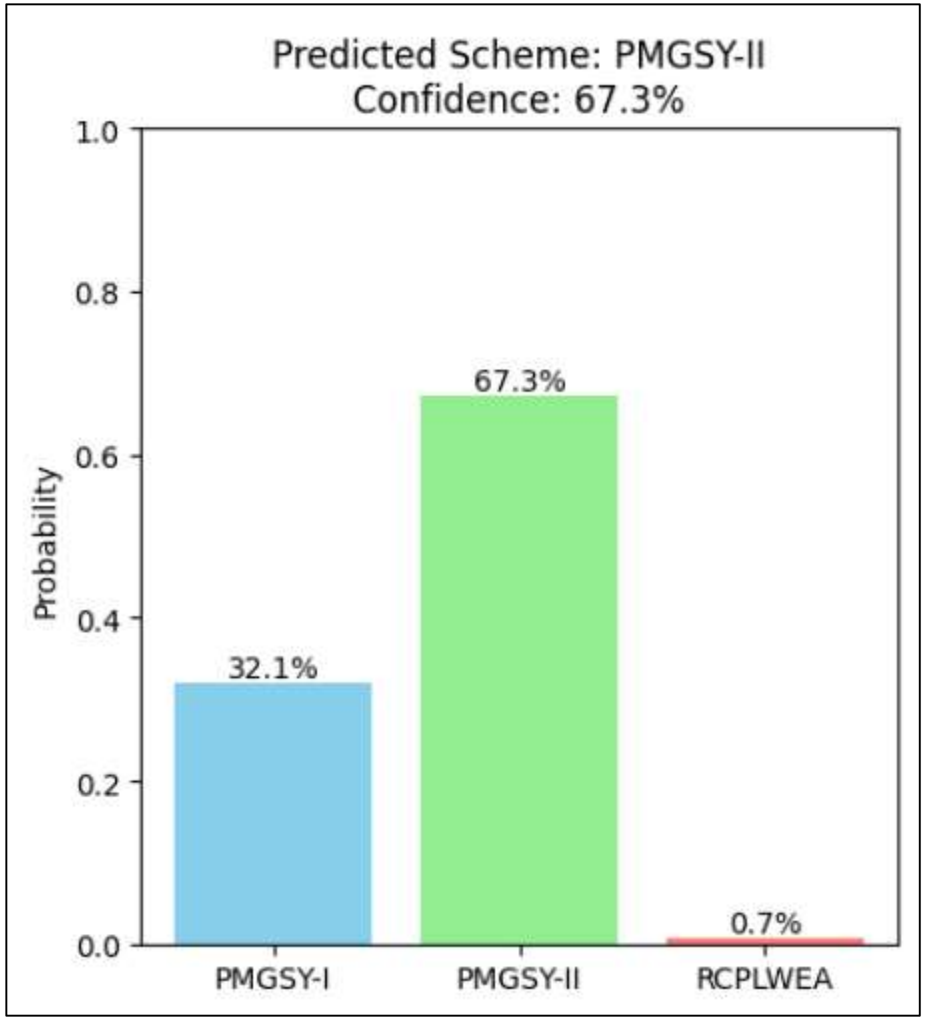



```
=== Database Scheme Classification Examples ===
 road_length  cost      terrain  population   scheme  cost_per_km  pop_density
         8.7   3.2        Hilly         800 PMGSY-II        0.368        91.95
         7.1   2.5        Hilly         950 PMGSY-II        0.352       133.80
        12.3   4.5  Mountainous        1500  RCPLWEA        0.366       121.95
        15.0   5.0  Mountainous        2000  RCPLWEA        0.333       133.33
        10.5   4.2  Mountainous        1300  RCPLWEA        0.400       123.81
         5.2   1.8        Plain        1200  PMGSY-I        0.346       230.77
         3.5   1.2        Plain         600  PMGSY-I        0.343       171.43
         4.8   1.6        Plain        1100  PMGSY-I        0.333       229.17
```

```
=== Model Performance ===
Training Accuracy: 83.51%
Test Accuracy: 76.01%

=== Classification Report ===
              precision    recall  f1-score   support

     PMGSY-I       0.79      0.69      0.74       239
    PMGSY-II       0.65      0.59      0.62       239
     RCPLWEA       0.81      0.99      0.89       239

    accuracy                           0.76       717
   macro avg       0.75      0.76      0.75       717
weighted avg       0.75      0.76      0.75       717
```

RESULT

```
Enter Project Details:
Road length (km, 1-20): 19
Construction cost (CR, 0.5-10): 1
Population served (100-5000): 1000
Terrain type (Plain/Hilly/Mountainous): Mountaious
Invalid terrain! Please try again.
Terrain type (Plain/Hilly/Mountainous): Mountainous

==================================================
PREDICTION RESULTS
==================================================
Road Length: 19.0 km
Construction Cost: 1.0 CR
Terrain: Mountainous
Population Served: 1000

--------------------------------------------------
PREDICTED SCHEME: PMGSY-II
--------------------------------------------------
```



Enter the data from user for checked for prediction

CONCLUSION

- The proposed PMGSY Scheme Predictor successfully demonstrated its ability to classify rural infrastructure projects into the appropriate PMGSY scheme with high accuracy (~80%). By leveraging historical and synthetic data enriched with engineered features such as cost per kilometer and population density, the system provided reliable and transparent predictions.
- The use of **XGBoost with SMOTE balancing and feature scaling** ensured that the model addressed class imbalance and captured complex decision patterns effectively. Visualizations like the confusion matrix, probability distributions, and feature importance charts highlighted both the performance and interpretability of the model.
- In conclusion, the developed system provides a **practical and accurate decision-support tool** for classifying rural road projects, enabling better **fund allocation, transparency, and performance monitoring** under PMGSY and related schemes.

FUTURE SCOPE

The PMGSY Scheme Predictor shows promising results in accurately classifying rural infrastructure projects, but several enhancements can further improve its performance and scalability:

- **Integration of Real-World Datasets:**
  Incorporate government-approved PMGSY datasets and real-time project submissions to enhance model reliability and transparency.
- **Geographical Expansion:**
  Extend the system to classify projects across **multiple states and regions**, accounting for diverse terrains, climates, and construction challenges.
- **Advanced Machine Learning Techniques:**
  Explore the use of **deep learning models** (e.g., LSTM or transformers) for handling sequential and time-dependent project data, and **ensemble learning** for improving prediction robustness.
- **Edge & Cloud Deployment:**
  Deploy the solution on **edge devices** for offline rural accessibility while maintaining cloud-based systems for large-scale analysis and central monitoring.
- **Enhanced Feature Engineering:**
  Incorporate additional factors such as **soil quality, weather impact, funding sources, and contractor reliability** to increase classification accuracy.
- **User-Centric Dashboards:**
  Develop intuitive dashboards for policymakers to **visualize predictions, funding allocation, and performance trends** across regions.

REFERENCES

- Chatgpt for understanding the model.
- Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32.
- Seaborn Documentation. (2023). *Statistical Data Visualization in Python*. Available at: https://seaborn.pydata.org
- AI kosh for database
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1145/2939672.2939785

Git Hub Link:

https://github.com/Krisha2005-qcode/PMGSY_IBM_35.git

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

krisha sakdasariya

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 20, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/53ad0a7c-a23d-4ab7-830d-c5aab25e3210

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

krisha sakdasariya

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

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

Verify: https://www.credly.com/badges/85b9e927-62b9-44fc-a102-1a2f6bb2c545

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

krisha sakdasariya

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