
CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:

MANOJIT SAHA

GARGI MEMORIAL INSTITUTE OF TECHNOLOGY

COMPUTER SCIENCE AND ENGINEERING

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 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.

We need 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 uses machine learning to automate the classification of rural infrastructure projects into their correct PMGSY scheme.

It involves:

- **Data Collection:** Download and preprocessed the data from AI Kosh Website.
- **Feature Engineering (FE):** From total of 14 columns, extracted PMGSY_SCHEME as our Label and other columns as our feature.
- **Model Training:** Use a Random Forest Classifier with two rounds of hyperparameter optimization (HPO1 and HPO2).
- **Deployment:** Use IBM Cloud Lite services to deploy the trained model and expose a prediction interface.

SYSTEM APPROACH

❑ System Requirements:

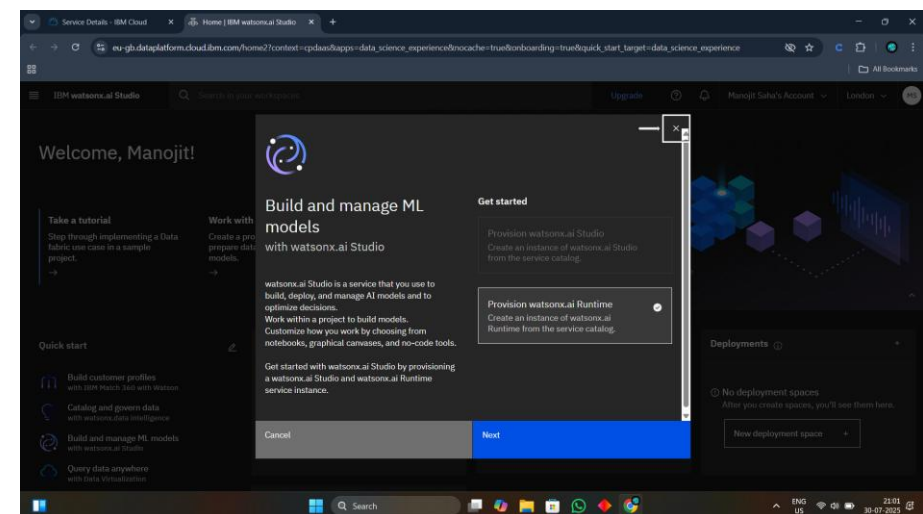
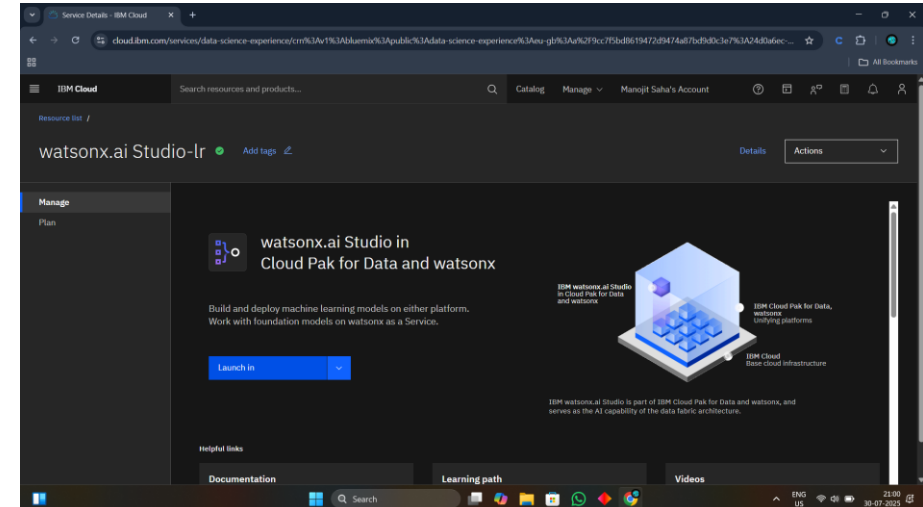
- IBM Cloud Lite (Watson Studio, Cloud Object Storage)
- Local/Cloud Jupyter Notebook
- IBM Watson Machine Learning for deployment

❑ Libraries Used :

- pandas, numpy – Data handling
- scikit-learn – Random Forest, HPO
- matplotlib, seaborn – Visualization
- ibm_watson_studio library– IBM Cloud access

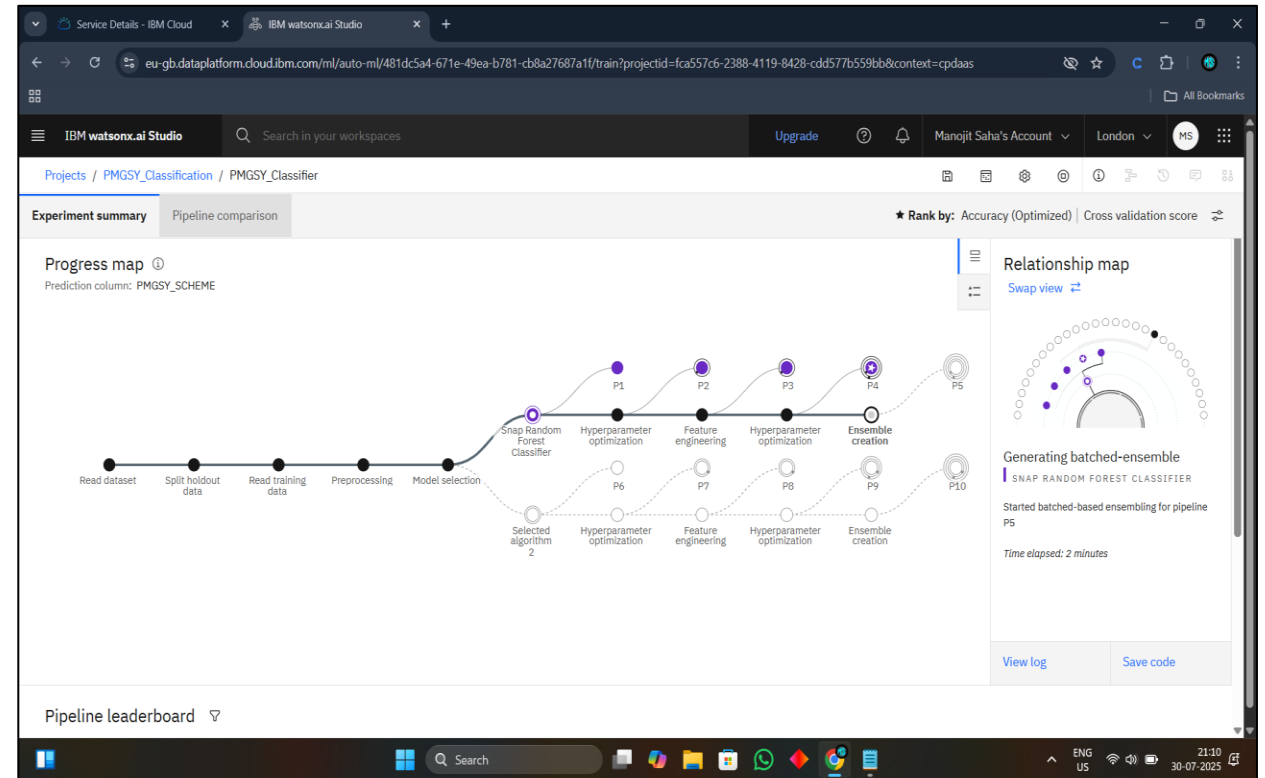
❑ Development Steps:

- Data upload to IBM Cloud Object Storage.
- Model Training: Random Forest with HPO1 → FE → HPO2
- Deployment via Watsonx.ai.studio as and endpoint.



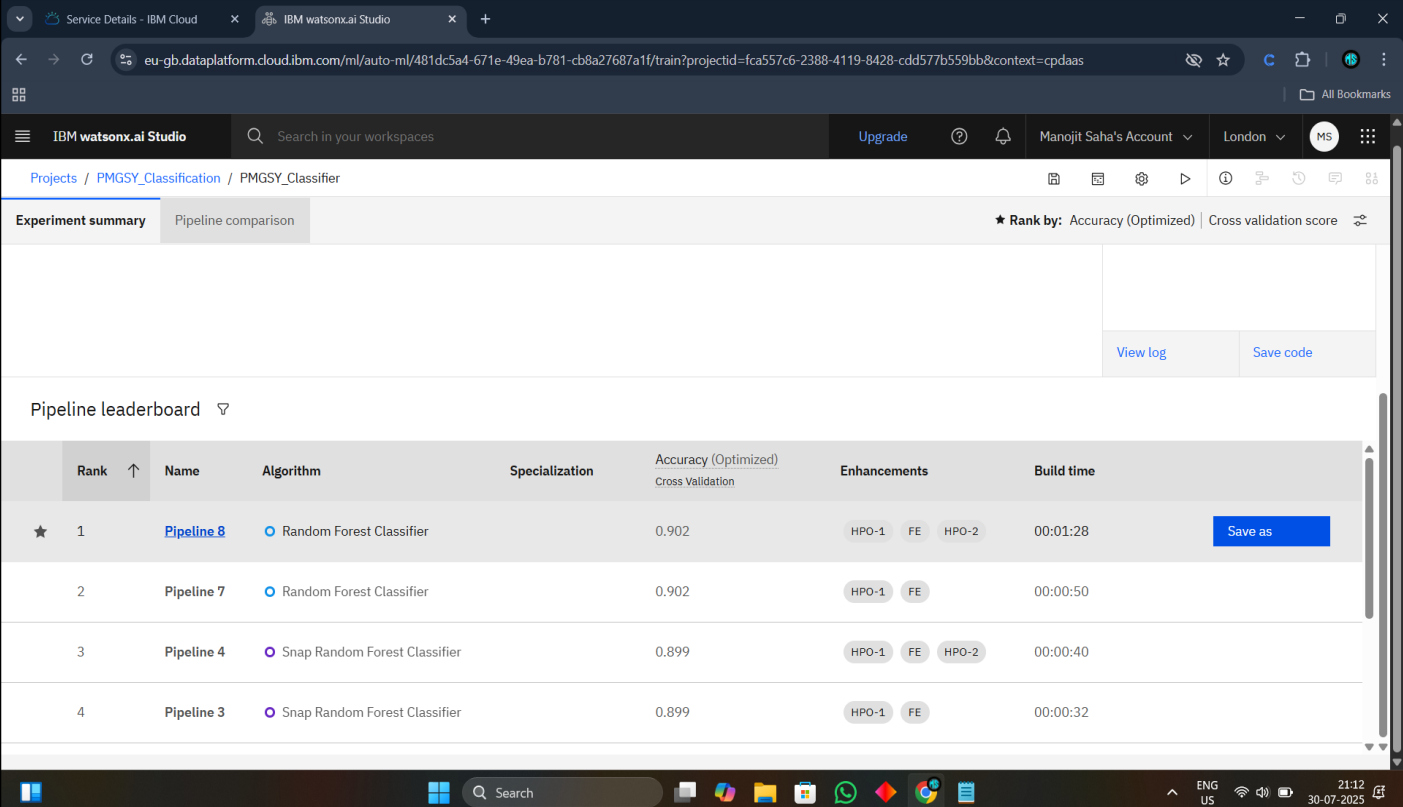
ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
 - Random Forest Classifier selected for its robustness and high performance on structured data.
- **Data Input:**
 - Total 14 features as input like Project length, cost, bridge/road type, funding details, etc.
 - One output PMGSY Scheme with a confidence score in percentage.
- **Training Process:**
 - Initial hyperparameter tuning → Feature selection & transformation → Final optimization for performance
 - Performance evaluated using accuracy, cross-validation score.
- **Prediction Process:**
 - Deploy on IBM Watson Machine Learning
 - Testing predictions via deployed endpoint in watsonx.ai.studio



RESULT

- After pipeline executes the best model is **Random Forest Classifier** algorithm.
- Model Accuracy:90%



The screenshot displays the IBM Watson AI Studio interface. The top navigation bar includes the IBM Watson AI Studio logo, a search bar, and user account information (Manojit Saha's Account, London). The main content area shows the 'Experiment summary' tab for the 'PMGSY_Classifier' project. The 'Rank by: Accuracy (Optimized)' is selected, and the 'Cross validation score' is visible. Below this, the 'Pipeline leaderboard' is shown, listing four pipelines. The first pipeline, 'Pipeline 8', is the top performer with an accuracy of 0.902 and a build time of 00:01:28. It is a 'Random Forest Classifier' algorithm. The other three pipelines (Pipeline 7, Pipeline 4, and Pipeline 3) also use the 'Random Forest Classifier' algorithm but have lower accuracies (0.899) and shorter build times (00:00:50, 00:00:40, and 00:00:32 respectively). The interface also includes buttons for 'View log' and 'Save code'.

Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★ 1	Pipeline 8	Random Forest Classifier		0.902	HPO-1 FE HPO-2	00:01:28 Save as
2	Pipeline 7	Random Forest Classifier		0.902	HPO-1 FE	00:00:50
3	Pipeline 4	Snap Random Forest Classifier		0.899	HPO-1 FE HPO-2	00:00:40
4	Pipeline 3	Snap Random Forest Classifier		0.899	HPO-1 FE	00:00:32

RESULT

- Predicted Result Based given Input
- Each prediction has it's own confidence percentage which represents how much model thinks it's prediction is correct.

The screenshot shows the IBM Watson AI Studio interface. The top navigation bar includes 'Service Details - IBM Cloud', 'PMGSY_scheme_classification', and a search bar. The main header shows 'Deployment spaces / PMGSY_Classification_Project / P8 - Random Forest Classifier: PMGSY_Classifier /'. The 'Test' tab is active, displaying 'Enter input data' with a 'Text' input field and a 'JSON' tab. Below the input field, there is a table with 5 columns: STATE_NAME (other), DISTRICT_NAME (other), NO_OF_ROAD_WORK_SANCTIONED (double), LENGTH_OF_ROAD_WORK_SANCTIONED (double), and NO_OF_BRIDGES_SANCTIONED (double). The table contains one row of data: West Bengal, Bankura, 2.2, 2.4, 4.5. A 'Predict' button is at the bottom right.

	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF_BRIDGES_SANCTIONED (double)
1	West Bengal	Bankura	2.2	2.4	4.5
2					
3					

The screenshot shows the 'Prediction results' modal window. It displays 'Prediction type: Multiclass classification' and 'Prediction percentage: 1 record'. A large purple circle with the number '1' and 'record' is shown. To the right, there is a table with 2 columns: Prediction and Confidence. The table contains one row: PMGSY-II, 50%. A 'Download JSON file' button is at the bottom right.

Prediction	Confidence
PMGSY-II	50%

CONCLUSION

- The Random Forest model with optimized hyperparameters effectively classifies infrastructure projects with 90% accuracy. The IBM Cloud deployment ensures scalable access for planners and policymakers, making it a powerful tool for project monitoring and analysis.

FUTURE SCOPE

- Integration with real-time project updates and GIS data
- Use of ensemble models or deep learning for better performance
- Expansion to other rural schemes beyond PMGSY
- Dashboard integration for government decision-makers

REFERENCES

- AI Kosh Dataset:
https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgysy.html
- IBM Cloud Documentation:
<https://cloud.ibm.com/docs>
- scikit-learn Documentation:
https://scikit-learn.org/stable/user_guide.html

IBM CERTIFICATIONS

- Getting started with AI



IBM CERTIFICATIONS

- Journey to Cloud: Envisioning Your Solution



IBM CERTIFICATIONS

- Retrieval Augmented Generation Lab





THANK YOU