
CAPSTONE PROJECT

PMGSY SCHEME CLASSIFIER

**Presented By:
Aakash Singh**

**IET, Dr. Shakuntala Misra National Rehabilitation University Lucknow
Department of Electronics and Communication Engineering**

OUTLINE

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

PROBLEM STATEMENT

In India, the Pradhan Mantri Gram Sadak Yojana (PMGSY) plays a vital role in connecting rural communities by building all-weather roads and bridges. Over time, this program has introduced multiple phases, each with its own goals and funding methods. With thousands of projects underway or completed, keeping track of which project belongs to which phase has become challenging. Doing this manually takes a lot of time, is prone to mistakes, and doesn't scale well. To support better planning, budgeting, and evaluation, we need an intelligent automated system that can quickly and accurately classify these projects based on their characteristics. This would help decision-makers monitor progress and make informed choices to improve rural connectivity.

PROPOSED SOLUTION

- The proposed system aims to streamline and automate the classification of road and bridge construction projects into their respective PMGSY schemes. This solution leverages data science and machine learning to accurately categorize projects based on physical and financial parameters. The main components are as follows:
- Data Collection:
 - Gather historical data on completed and ongoing road and bridge projects, including their physical specifications, financial details, project status, and existing scheme labels.
 - Source additional relevant datasets, such as government records or the AI Kosh dataset, to ensure diversity and accuracy in the training data.
- Data Preprocessing:
 - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
 - Perform feature engineering to identify and construct the most impactful features for distinguishing between different PMGSY schemes.
- Machine Learning Algorithm:
 - Utilize IBM AutoAI to automate the selection, training, and optimization of machine learning classification models suited to this problem.
 - Let the model learn from financial, physical, and administrative features to predict the appropriate scheme for each project.
- Deployment:
 - Deploy the trained model on IBM Cloud Lite services, ensuring scalability and accessibility for end-users, planners, and policy makers.
 - Integrate the solution into an interface or API, allowing stakeholders to quickly classify new projects as data becomes available.
- Evaluation:
 - Assess model performance using relevant metrics such as accuracy, precision, recall, and the confusion matrix.
 - Result : The system now enables fast, reliable, and accurate classification of rural infrastructure projects into their correct PMGSY schemes, reducing manual workload and supporting better government decision-making.

SYSTEM APPROACH

- **Platform:** IBM Cloud Lite services for scalable, cloud-based model development and deployment.
- **Tool:** IBM AutoAI for automated machine learning pipeline including data preprocessing, model training, and optimization.
- **Data Source:** AI Kosh dataset from the Government of India containing detailed road and bridge project data.
- **Programming Languages:** Python (for model exploration, evaluation, and deployment scripting).
- **Model Deployment:** Deployed trained model as an API endpoint on IBM Cloud for real-time project classification.
- **Additional Tools:** Jupyter notebooks for data analysis and visualization during development.

ALGORITHM & DEPLOYMENT

- **Algorithm:**

Utilized IBM AutoAI for automated machine learning, which conducted feature selection, model training, optimization, and pipeline generation. The approach involved multiclass classification to predict the PMGSY scheme based on project features.

- **Model Training:**

The AutoAI experiment was configured using the AI Kosh dataset, targeting the "PMGSY_SCHEME" column for classification. Training included holdout validation and scoring using accuracy.

- **Model Evaluation:**

Pipelines were compared using accuracy and feature importance; the best pipeline was selected for deployment.

- **Deployment:**

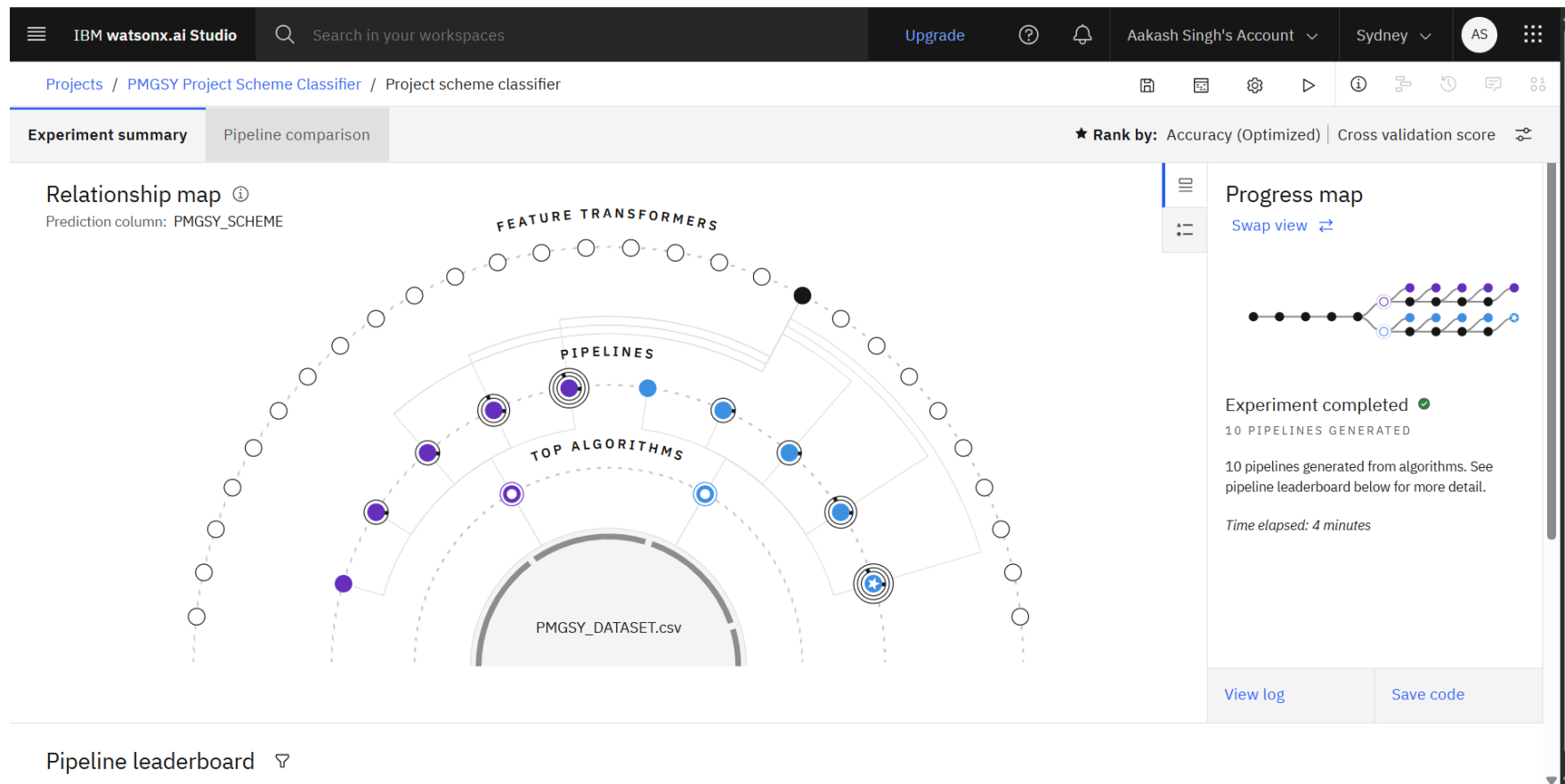
The final model was deployed on IBM Cloud Lite as a web service using IBM Watsonx.ai Runtime. It is accessible via API for real-time project classification.

- **Scoring & Prediction:**

The deployed model accepts new project data, returning predicted PMGSY scheme along with probability scores, enabling accurate and quick scheme assignment.

RESULT

- **Model Performance:** The top pipeline, using a Batched Tree Ensemble Classifier (XGB Classifier), achieved a cross-validated accuracy of 92.4%, making it highly reliable for classifying PMGSY project schemes.



RESULT

- **Pipeline Development:** AutoAI generated 10 different pipelines through data preprocessing, model selection, feature engineering, and ensemble methods. Pipeline 10 was selected as the best performer based on accuracy.

The screenshot displays the IBM watsonx.ai Studio interface. The top navigation bar includes the IBM watsonx.ai Studio logo, a search bar, an Upgrade button, and user account information (Aakash Singh's Account, Sydney, AS). The breadcrumb trail shows the path: Projects / PMGSY Project Scheme Classifier / Project scheme classifier. The main content area is divided into two tabs: Experiment summary (selected) and Pipeline comparison. The Experiment summary tab shows a progress bar with stages: optimization, engineering, optimization, and Creation. A summary box indicates that 10 pipelines were generated from algorithms, with a time elapsed of 4 minutes. Below this, the Pipeline leaderboard is displayed, showing a table of generated pipelines ranked by Accuracy (Optimized) Cross Validation score.

	Rank ↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 10	Batched Tree Ensemble Classifier (XGB Classifier)	INCR	0.924	HPO-1 FE HPO-2 BATCH	00:01:59
	2	Pipeline 9	XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:55
	3	Pipeline 8	XGB Classifier		0.924	HPO-1 FE	00:01:17
	4	Pipeline 7	XGB Classifier		0.918	HPO-1	00:00:21

RESULT

- **Deployment & Usage:** The final model was seamlessly deployed as an online service on IBM Cloud. Users can input new project data directly into the interface to receive instant scheme predictions.

The screenshot displays the IBM watsonx.ai Studio interface. At the top, the navigation bar includes the IBM watsonx.ai Studio logo, a search bar, an 'Upgrade' button, a help icon, a notification bell with a red '1', and user account information for 'Aakash Singh's Account' with a dropdown menu, 'Sydney' with a dropdown, and a profile icon 'AS' with a settings menu. Below the navigation bar, the breadcrumb trail reads 'Deployment spaces / Project 1 / XGB Classifier: Project scheme classifier'. The main content area is titled 'PMGSY: Project Scheme' and shows a 'Deployed' status with a green checkmark and an 'Online' label. Underneath, there are tabs for 'API reference' and 'Test'. The 'Test' tab is active, showing an 'Enter input data' section. This section has two tabs: 'Text' (selected) and 'JSON'. Below the tabs, a message states 'Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.' There are three links: 'Download CSV template' (with a download icon), 'Browse local files' (with an external link icon), and 'Search in space' (with an external link icon). A 'Clear all' link with an 'x' icon is also present. A table is displayed with 5 rows and 6 columns. The columns are: '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 rows contain data for Jammu And Kashmir, Jharkhand, Ladakh, Madhya Pradesh, and Telangana. A 'Predict' button is located at the bottom right of the table area.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Aakash Singh's Account

Sydney

AS

Deployment spaces / Project 1 / XGB Classifier: Project scheme classifier

PMGSY: Project Scheme

Deployed Online

API reference

Test

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

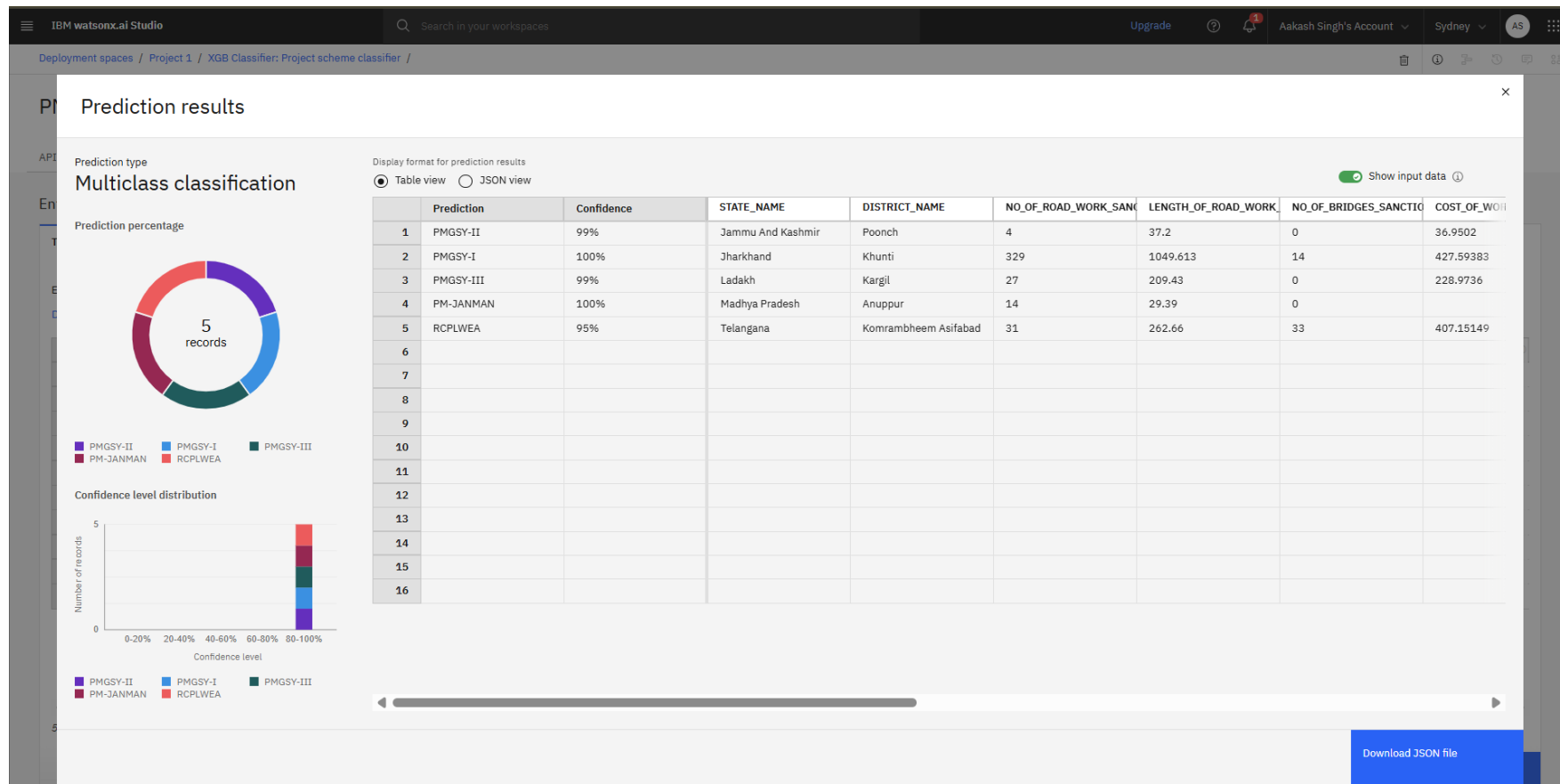
	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF_BRIDGES_SANCTIONED (double)
1	Jammu And Kashmir	Poonch	4	37.2	0
2	Jharkhand	Khunti	329	1049.613	14
3	Ladakh	Kargil	27	209.43	0
4	Madhya Pradesh	Anuppur	14	29.39	0
5	Telangana	Komrambheem Asifabad	31	262.66	33
6					

5 rows, 14 columns

Predict

RESULT

- **Prediction Results:** On test data, the classifier confidently assigned schemes such as PMGSY-II, PMGSY-I, PMGSY-III, PM-JANMAN, and RCPLWEA with confidence levels between 95% and 100%. The system displayed both predicted classes and their associated probabilities for transparency.



CONCLUSION

- This project successfully automated the classification of rural road and bridge projects into relevant PMGSY schemes using IBM AutoAI and cloud deployment. The resulting model, with a high accuracy of 92.4%, enables government planners and stakeholders to make quick and reliable decisions while eliminating the errors, delays, and resource constraints of manual classification. Our solution demonstrates the impact of machine learning in public sector infrastructure management—boosting transparency, driving efficiency, and supporting scalable policy analysis. Through rigorous model evaluation and seamless integration, the system stands ready to streamline project monitoring and future scheme assessments.

FUTURE SCOPE

- **Expanded Data Integration:** Incorporate real-time project updates and additional data sources, such as satellite imagery and local weather, to further improve classification accuracy and adaptability.
- **Enhanced Algorithmic Approaches:** Explore advanced techniques (e.g., deep learning, ensemble meta-models) for even more robust project classification.
- **Geographical & Temporal Analysis:** Add modules for region-wise scheme trends, project progress over time, and predictive analytics for upcoming project needs.
- **Policy & Impact Assessment:** Use the system for deeper evaluation of scheme effectiveness, impact measurement, and data-driven policy recommendations.
- **User Experience:** Develop a more intuitive dashboard and APIs to allow government officials and planners to upload, visualize, and analyze project data easily.
- **Scalability:** Extend the solution framework to cover other government schemes or public infrastructure domains using similar AI-based classification methods.

REFERENCES

- AI Kosh Dataset: Pradhan Mantri Gram Sadak Yojana (PMGSY)
 - https://aikosh.indiaai.gov.in/home/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html
- IBM Watson Studio AutoAI Documentation
 - <https://www.ibm.com/docs/en/watsonx/w-and-w/2.0.0?topic=models-autoai>
- IBM watsonx.ai API Reference
 - <https://cloud.ibm.com/apidocs/watsonx-ai>
- PMGSY Official Guidelines and Publications
 - <https://www.myscheme.gov.in/schemes/pmgsy>
 - <https://megcnrd.gov.in/forms/PMGSY.pdf>
- Machine Learning Classification Best Practices
 - <https://www.geeksforgeeks.org/machine-learning/getting-started-with-classification/>
 - <https://www.ibm.com/docs/en/watsonx/w-and-w/2.0.0?topic=models-build-model-autoai>

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Aakash Singh

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/96dd8b4c-f90e-4aa2-868f-41a79e8d2c66>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Aakash Singh

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/b90ad23c-a4f7-42d8-8e65-ef8589c92b38>



IBM CERTIFICATIONS

IBM SkillsBuild

Completion Certificate



This certificate is presented to

Aakash Singh

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