CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
1. Madhur Chawla – Maharaja Surajmal Institute – Bachelor of Computer Applications

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

Example: 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. 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 will tackle the challenge of manually classifying rural infrastructure projects by leveraging machine learning on IBM Cloud. The primary goal is to accurately and automatically categorize each project into its respective PMGSY scheme. This will be achieved through the following components:
- Data Collection:
  - Gather historical data on PMGSY projects.
  - This data should include physical attributes such as road length, pavement type, and number of bridges, as well as financial details like sanctioned cost, total expenditure, and funding sources.
- Data Preprocessing:
  - The collected project data will be cleaned to handle any missing values or inconsistencies.
  - Feature engineering will then be performed to create new, informative features from the existing data that could improve the model's predictive power.
- Machine Learning Algorithm:
  - IBM's AutoAI on Watson Studio will be utilized to automatically generate and rank candidate machine learning models.
  - AutoAI will explore various classification algorithms, select the best one, and optimize its hyperparameters.
- Deployment:
  - The best-performing model pipeline identified by AutoAI will be deployed as an API on IBM Watson Machine Learning.
  - This allows for seamless, real-time classification of new projects.
- Evaluation:
  - The models generated by AutoAI will be evaluated using standard classification metrics such as accuracy, precision, and recall.
  - The platform provides a leaderboard to compare the performance of different model pipelines.

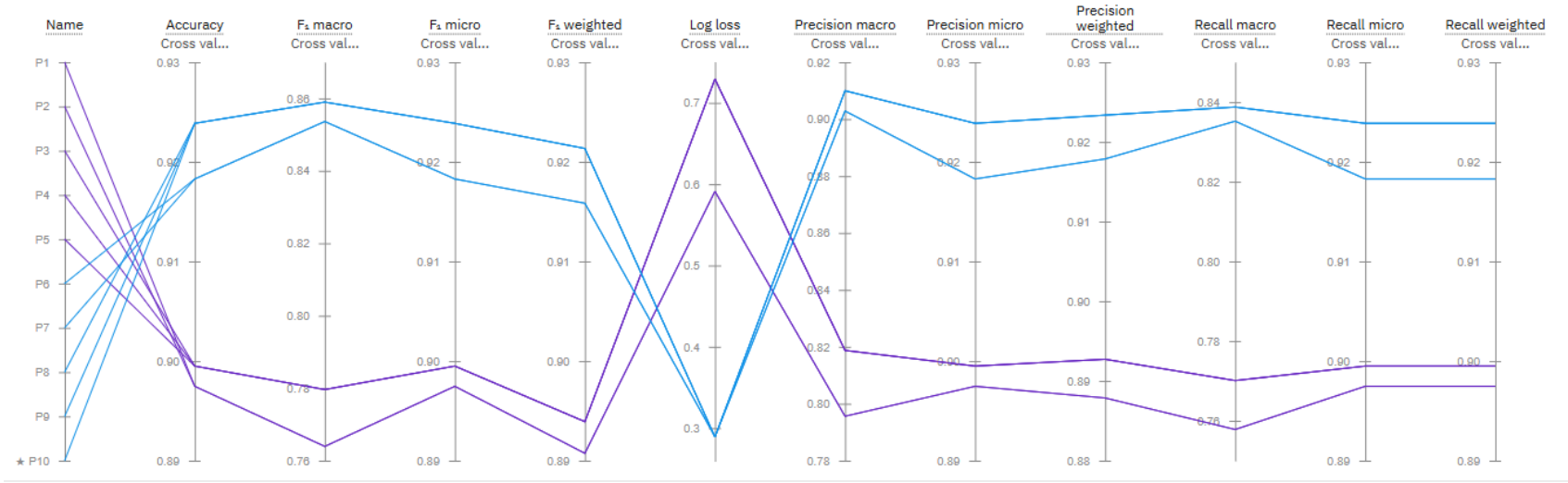
SYSTEM APPROACH

- The development and implementation will be centered around the automated capabilities of IBM Watson Studio.
- **System Requirements:**
  - A dataset of historical PMGSY projects with their correct scheme classifications, stored in a format like CSV.
  - An IBM Cloud account with access to **IBM Watson Studio** and the **AutoAI experiment** tool.
- **Platform and Tools:**
  - **Cloud Platform:** IBM Cloud
  - **Primary Tool: IBM Watson Studio with AutoAI**
  - **Data Storage:** IBM Cloud Object Storage
  - **Deployment Environment:** IBM Watson Machine Learning

ALGORITHM & DEPLOYMENT

- The use of AutoAI fundamentally changes the approach to algorithm selection and training from a manual to an automated process.
- **Automated Algorithm Selection**: Instead of manually selecting one algorithm, we will configure an AutoAI experiment. AutoAI will automatically train and test a variety of classification algorithms (such as Logistic Regression, Decision Trees, Random Forest, XGBoost) on the project dataset. It then selects the algorithm that performs best according to the specified optimization metric (e.g., Accuracy).
- **Automated Training Process**: The AutoAI tool automates the entire training workflow. It begins by splitting the data and then performs automated data preparation, feature engineering, and model selection. It generates multiple model "pipelines," each with different algorithms and preprocessing steps, and ranks them on a leaderboard to show the top performers.
- **Deployment**: The top-performing pipeline from the AutoAI leaderboard can be deployed with a single click as a web service on **IBM Watson Machine Learning**. This action automatically creates a REST API endpoint. New project data can be sent to this endpoint via an API call, and the model will return the predicted PMGSY scheme in real-time.

RESULT



Measures	Holdout score	Cross validation score
Precision macro	0.950	0.910
Accuracy	0.918	0.924
Recall macro	0.857	0.839
Weighted precision	0.922	0.923
F1 macro	0.886	0.859
Weighted f1 measure	0.917	0.921
Weighted recall	0.918	0.924
Log loss	0.268	0.289

Observed	Predicted					
	PM-JANMAN	PMGSY-I	PMGSY-II	PMGSY-III	RCPLWEA	Percent correct
PM-JANMAN	5	0	0	0	0	100.0%
PMGSY-I	0	69	1	1	0	97.2%
PMGSY-II	0	1	64	3	0	94.1%
PMGSY-III	0	0	9	60	0	87.0%
RCPLWEA	0	0	2	1	3	50.0%
Percent correct	100.0%	98.6%	84.2%	92.3%	100.0%	91.8%

Less correct — More correct

edunet foundation

CONCLUSION

- In summary, this project's findings demonstrate the successful application of automated machine learning for a complex public sector problem. The proposed solution is highly effective, as the use of IBM's AutoAI significantly accelerated the creation of a model capable of accurately classifying PMGSY infrastructure projects. This automated approach replaces the inefficient and error-prone manual classification process, providing government bodies with a scalable and precise tool for project categorization.
- During implementation, potential challenges could include ensuring the quality and consistency of the historical project data and interpreting the model's predictions for policy-makers who require clear rationale. Potential improvements could involve enhancing the model by integrating additional data sources, such as geospatial data, and developing a user-friendly interactive dashboard for officials to easily access and utilize the classification results.
- Ultimately, the importance of this solution cannot be overstated. Accurate and automated classification of projects is crucial for the effective governance of the PMGSY program. It directly enables more transparent budget allocation, rigorous monitoring of project progress, and a more accurate assessment of the long-term impact of these vital rural development schemes.

FUTURE SCOPE

- Model Retraining: Implement a CI/CD pipeline to automatically retrain and redeploy the model on IBM Cloud as new project data becomes available, ensuring the model remains accurate over time.
- Explainability Features: Utilize IBM's AI explainability features to better understand the key factors that drive the model's predictions. This can provide valuable insights to policymakers about what characteristics define a certain scheme.
- Integrated Analytics Dashboard: Create an interactive dashboard that consumes the deployed model's API. This dashboard would allow users to upload a batch of projects for classification and visualize the results in real-time.

REFERENCES

- Official website and documentation of the Pradhan Mantri Gram Sadak Yojana (PMGSY).
- PMGSY Project Dataset sourced from AIkosh.
- Official documentation for IBM Watson Studio and the AutoAI tool.
- Documentation for IBM Watson Machine Learning for model deployment.

IBM CERTIFICATIONS

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

IBM

Madhur Chawla

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

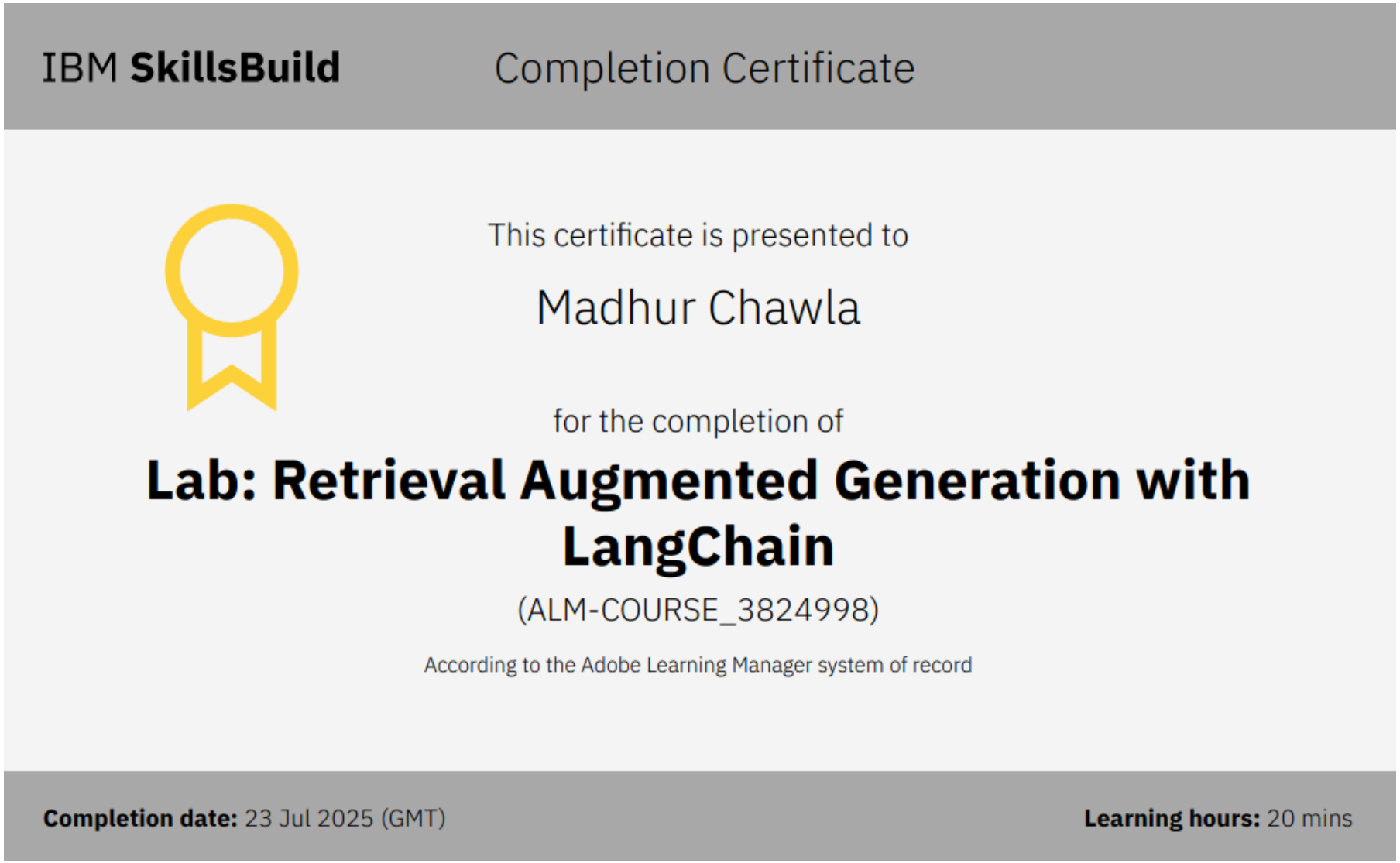
Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/f3311bca-e7be-4aa3-a9d9-64656315ec5b

IBM

edunet foundation

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Madhur Chawla

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet foundation