CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
Tejashwini S, MCA, REVA University

OUTLINE

- **Problem Statement** (Should not include solution)

- **Proposed System/Solution**

- **System Development Approach** (Technology Used)

- **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 with various distinct schemes. It is important for government bodies and planners to efficiently categorize ongoing and completed projects for effective monitoring and budget allocation. The crucial part is to automatically classify a road or bridge project into its correct PMGSY scheme based on its physical and financial characteristics.

PROPOSED SOLUTION

The project aims to address the challenge of automatically classifying rural infrastructure projects into their correct PMGSY scheme. The model will predict the appropriate scheme (e.g., PMGSY-I, PMGSY-II, RCPLWEA, etc.) for a project based on its physical and financial characteristics. The solution was built and deployed using IBM Cloud's watsonx.ai Studio and consists of the following components:

- **Data Collection**:

  - The project utilizes a dataset named PMGSY_DATASET.csv from AI Kosh Datasets.

  - The dataset has 15 columns and a size of 170.49 KB.

  - It includes features such as STATE_NAME, DISTRICT_NAME, NO_OF_ROAD_WORK_SANCTIONED, and LENGTH_OF_ROAD_WORK_SANCTIONED.

- **Data Preprocessing**:

  - The process uses an AutoAI experiment, which handles preprocessing automatically.

  - The dataset is split into 90% for training data and 10% for holdout data.

- **Machine Learning Algorithm**:

  - The model's objective is to predict the PMGSY_SCHEME.

  - The prediction type is Multiclass Classification.

  - The AutoAI experiment generated 10 pipelines from different algorithms.

  - The top-ranked pipeline (Pipeline 10) uses a Batched Tree Ensemble Classifier (XGB Classifier).

edunet
foundation

PROPOSED SOLUTION - CONTINUED

- **Deployment:**

  - **A deployment space named PMGSY_Project was created to deploy the model.**

  - **The top-performing model (P10 - XGB Classifier: PMGSY_Project) was promoted to this deployment space.**

  - **The model was deployed as an Online deployment, which runs the model in real-time.**

  - **The deployment provides public and private endpoints for scoring**

- **Evaluation:**

  - **The models were evaluated and ranked based on Accuracy (Optimized).**

  - **The top-ranked pipeline, Pipeline 10, achieved a cross-validation accuracy of 0.924.**

  - **The experiment also provides metrics such as F1 macro, precision, and recall for pipeline comparison.**

- **Result:**

  - **The model was tested with input data including STATE_NAME as Karnataka and DISTRICT_NAME as Bangalore.**

  - **The prediction result for this input was PMGSY-III**

edunet
foundation

PROPOSED SOLUTION - CONTINUED

- **<u>Benefits</u>:**

  - This intelligent classification will help government bodies and planners in monitoring, budgeting, and assessing the impact of schemes.

  - It provides a more efficient, less error-prone, and scalable alternative to manual classification methods.

SYSTEM APPROACH

System and Library Requirements for the PMGSY Project Classification Model

- The proposed solution was built using a combination of a cloud-based platform and machine learning.

- System requirements:

- Hardware:

  - CPU: The AutoAI experiment was configured to run in an environment with 8 CPU cores.

  - RAM: The configured environment for the experiment had 32 GB RAM.

- Software:

  - Platform: The entire project was developed within IBM watsonx.ai Studio.

  - Runtime: The models and notebooks use the watsonx.ai Runtime. The auto-generated notebook specifies Python 3.11 as part of the runtime environment.

edunet
foundation

SYSTEM APPROACH - CONTINUED

Library required to build the model:

- <u>Data Processing and Analysis</u>:

  - The project uses AutoAI, which automatically handles data processing.

  - The notebook generated by AutoAI is based on a scikit-learn pipeline.

- <u>Machine Learning Models</u>:

  - Scikit-learn: The auto-generated notebook for training continuation references scikit-learn 1.3.XGB Classifier:

  - The top-performing models in the pipeline leaderboard include XGB Classifier and Batched Tree Ensemble Classifier (XGB Classifier).

- <u>Model Persistence</u>:

  - The final model is saved as an asset and then promoted to a deployment space for future use.

ALGORITHM & DEPLOYMENT

- <u>Data Import and Pre-processing</u>

  - The project begins with uploading the PMGSY_DATASET.csv file into the IBM watsonx.ai Studio environment.

  - The data is then processed automatically using an AutoAI experiment.

  - This automated process handles tasks like feature engineering and preprocessing without manual intervention.

- <u>Train-Test Split</u>

  - The AutoAI tool automatically splits the data into a training set and a holdout set.

  - The split consists of 90% of the data for training and 10% for holdout data.

- <u>Model Training and Evaluation</u>

  - The AutoAI experiment automatically trains and evaluates multiple machine learning models.

  - Algorithms like XGB Classifier and Batched Tree Ensemble Classifier were applied.

  - The models were ranked on a "Pipeline leaderboard" based on their accuracy score.

edunet
foundation

ALGORITHM & DEPLOYMENT - CONTINUED

- **Comparison and Selection**
  - The "Pipeline leaderboard" compares the accuracy scores of all the generated models.
  - The top-ranked model, "Pipeline 10," uses a Batched Tree Ensemble Classifier and was selected for its performance.

- **New Data Prediction**
  - The deployed model provides an online interface under the "Test" tab to enter new data.
  - For a new data point with STATE_NAME as "Karnataka" and DISTRICT_NAME as "Bangalore," the model can predict the PMGSY scheme.

- **Model Persistence**
  - The best-performing model (Pipeline 10) is promoted to a deployment space.
  - The model is then deployed as an online service, making it accessible via public and private endpoints for future use.

- **GUI Integration**
  - The entire process, including data uploading, model building, and testing, is conducted through the web-based graphical user interface of IBM watsonx.ai Studio.
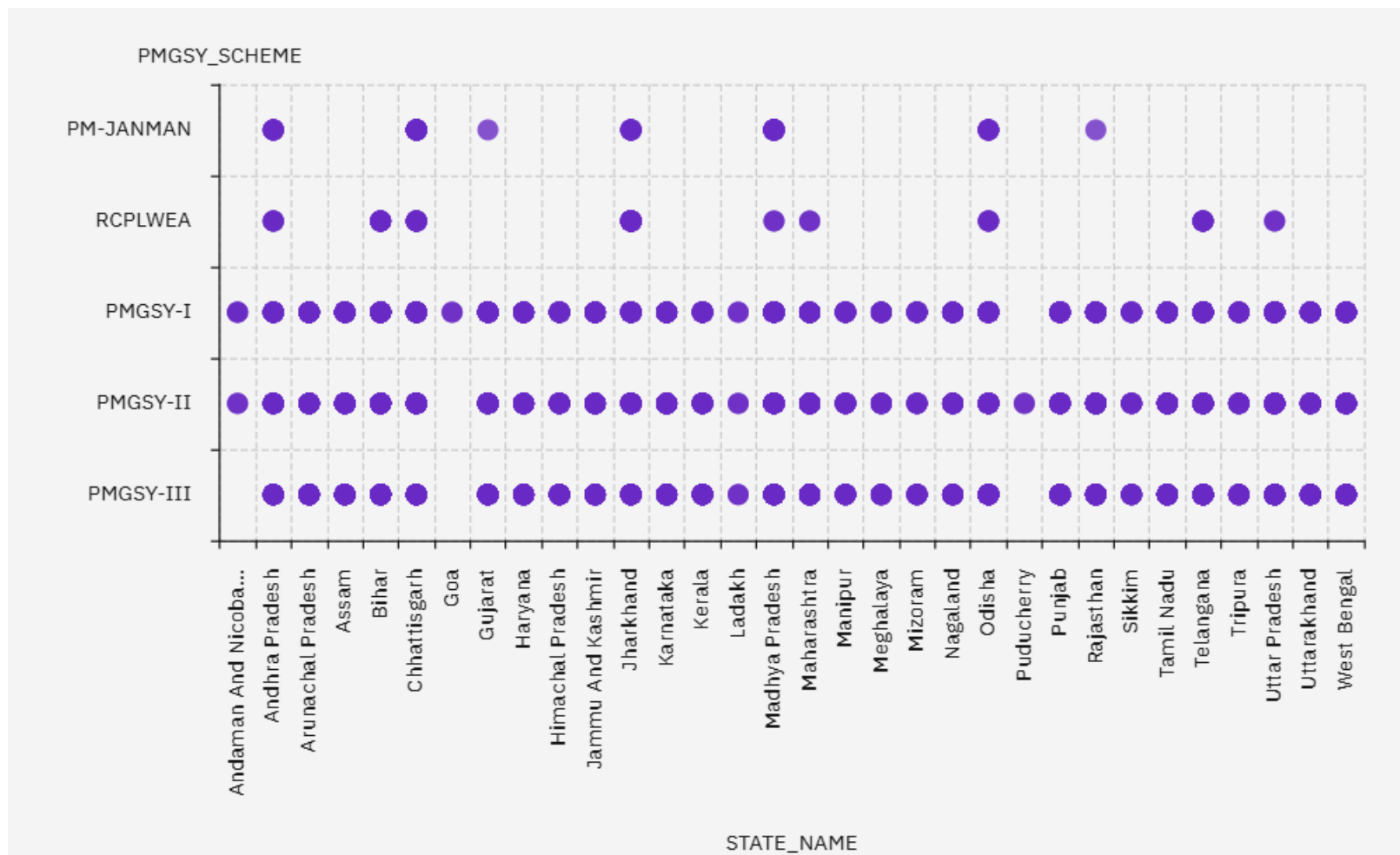
- **Result:**
  - The best-performing model achieved a cross-validation accuracy of 0.924.For the test input, the model successfully predicted the scheme as PMGSY-III.
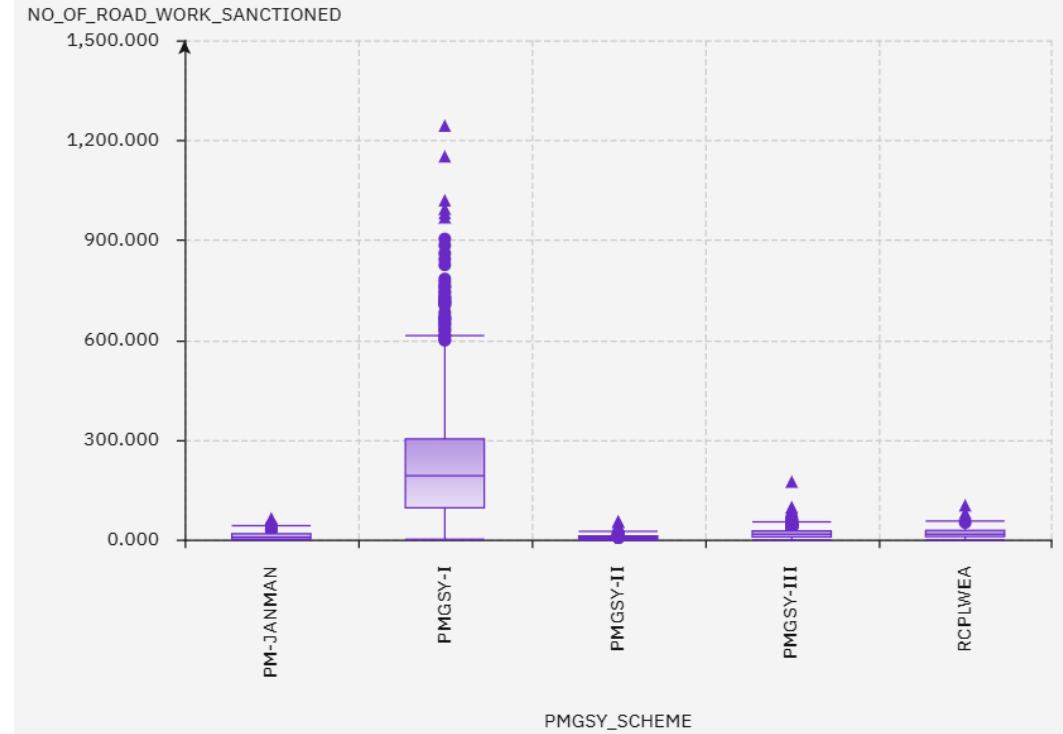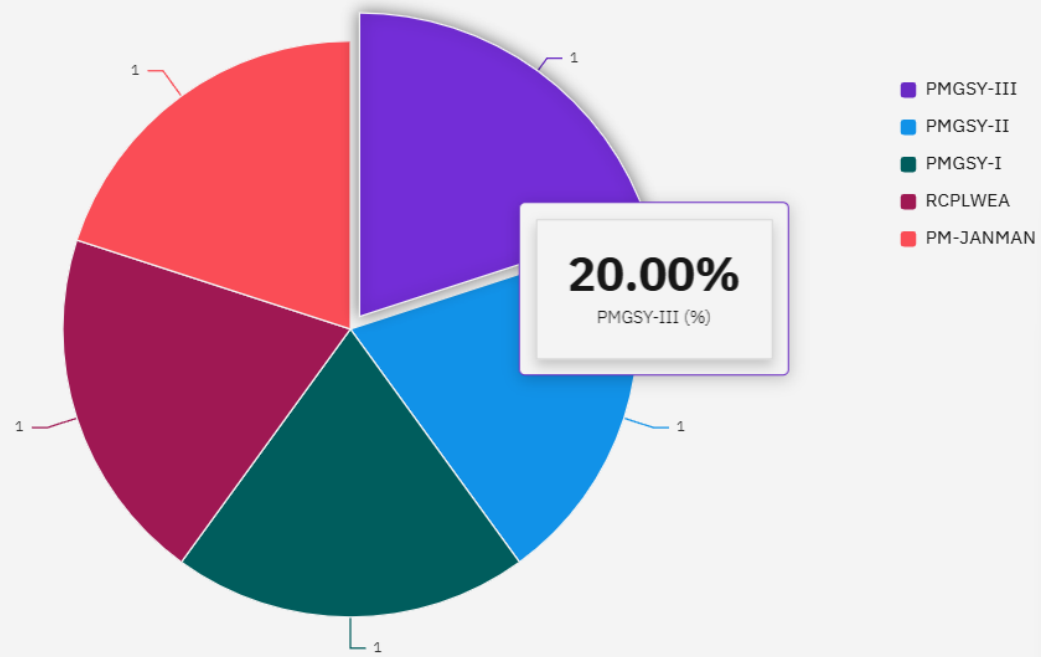
RESULT

The best-performing model achieved a cross-validation accuracy of **0.924**. The project successfully built a machine learning model to predict the PMGSY scheme of a project. When tested with specific data for Karnataka, the model predicted the scheme as **PMGSY-III**. This intelligent classification assists government bodies in monitoring and budget allocation by providing a reliable and automated solution.
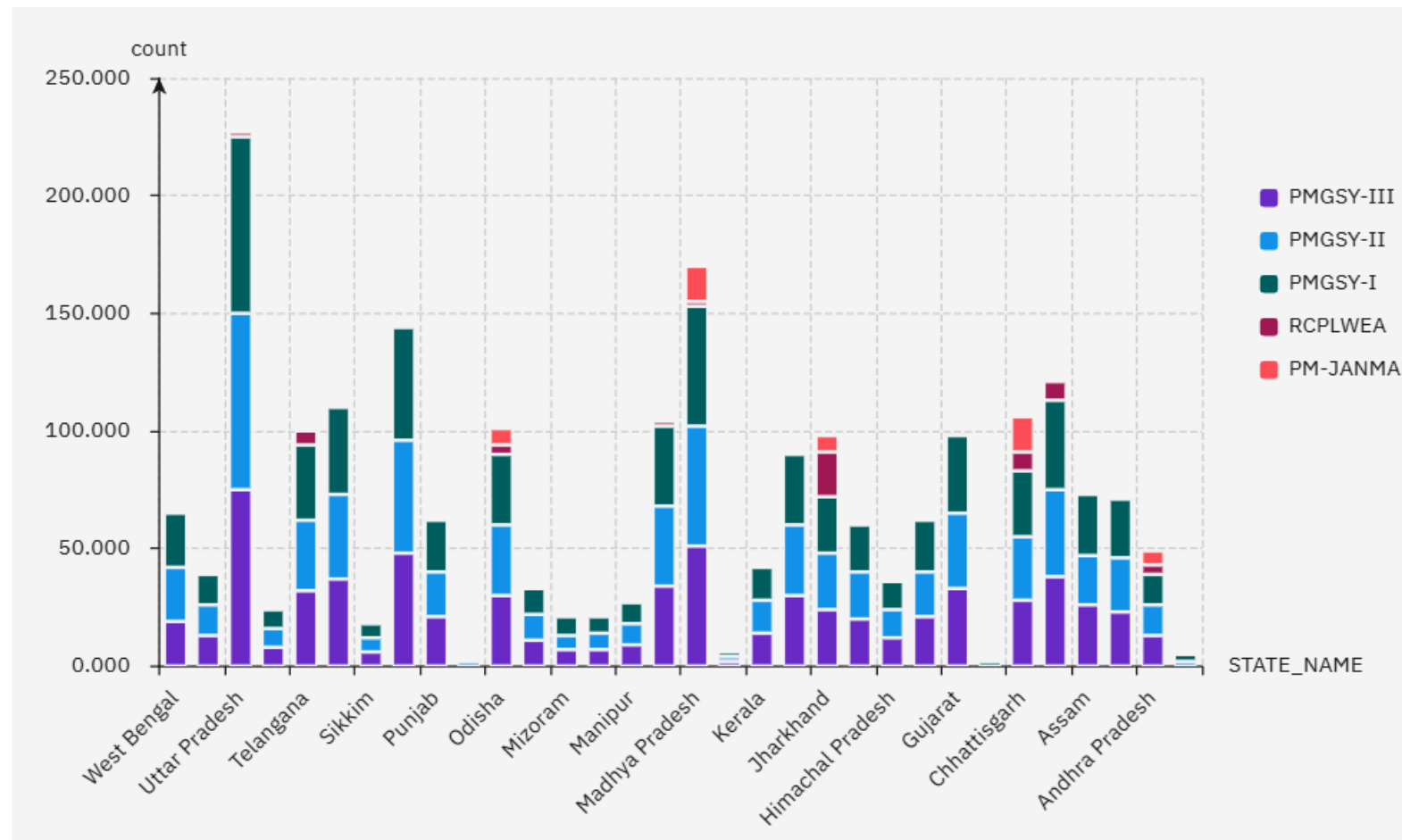
RESULT

RESULT

RESULT

CONCLUSION

- The project successfully built and deployed a machine learning model using IBM watsonx.ai Studio. The model's primary goal is to automatically classify rural infrastructure projects into their correct PMGSY scheme. The best-performing model, a Batched Tree Ensemble Classifier, achieved a cross-validation accuracy of **0.924**. This intelligent classification provides a solution for government bodies to move away from manual, time-consuming methods, thereby assisting in monitoring and budget allocation.

FUTURE SCOPE

- The model's performance could be further enhanced by incorporating a wider range of data sources, such as geospatial data or textual project descriptions.

- The deployed solution can be integrated into a larger analytics dashboard for real-time monitoring of project classification across different regions.

- The system could be extended to include a feedback loop, allowing for continuous retraining of the model with new project data.

- A user-friendly front-end application could be developed to allow non-technical users to input project details and receive instant classification predictions.

REFERENCES

- Dataset:
  https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

- IBM Cloud: https://cloud.ibm.com/

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Tejashwini S

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/b3fc84af-bb41-43dd-95ea-338a4dbdace6

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Tejashwini S

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/adf155c9-fef1-41bb-9a98-cd3dfed6f661

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild**   Completion Certificate

This certificate is presented to

Tejashwini S

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 25 Jul 2025 (GMT)   **Learning hours:** 20 mins

THANK YOU