CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS USING MACHINE LEARNING

Presented By:
Student Name-Rishita Singh
College Name- Greater Noida College
Department- Computer Science and Engineering

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

The Pradhan Mantri Gram Sadak Yojana (PMGSY) aims to enhance rural connectivity through road and bridge projects. Over time, it has expanded into multiple schemes like PMGSY-I, PMGSY-II, and RCPLWEA, each with unique characteristics.

Manual classification of thousands of ongoing and completed infrastructure projects into the appropriate scheme is inefficient, error-prone, and non-scalable. This creates bottlenecks in monitoring, funding, and policy evaluation.

PROPOSED SOLUTION

The proposed system automates the classification of rural infrastructure projects into PMGSY schemes using machine learning. It includes:

- Data Collection:
  Gather data on project type, length, cost, duration, terrain, location, and funding agency.
  Include temporal features like year of initiation and scheme timelines.

- Data Preprocessing:
  Handle missing values, encode categories, and normalize data.
  Apply feature engineering to improve model accuracy.

- Machine Learning Algorithm:
  Use classification algorithms (e.g., Random Forest, XGBoost).
  Train and tune the model using historical labeled project data.

- Deployment:
  Build an interface for real-time scheme prediction.
  Deploy the model on IBM Cloud using Watson Machine Learning.

- Evaluation:
  Evaluate with metrics like accuracy and F1-score.
  Monitor and retrain the model for continuous improvement.

SYSTEM APPROACH

- **System requirements**

  Platform & Tools:

  IBM Watson Studio

  IBM Cloud Object Storage

  IBM Cloud Machine Learning Service

  Python, Jupyter Notebook

  scikit-learn, pandas, matplotlib

- **Library required to build the model**

  `pandas`, `numpy` – Data manipulation

  `scikit-learn` – ML algorithms

  `seaborn`, `matplotlib` – Visualization

  `joblib` – Model serialization

  IBM Watson ML SDK – Model deployment

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

The model used is an **XGBoost Classifier** with a Batched Tree Ensemble specialization, selected for its high accuracy and ability to handle structured tabular data. It outperformed other models in AutoAI pipelines, achieving an accuracy of **92.4%**.

- **Data Input:**

The model uses the following input features:

- LENGTH_OF_ROAD_WORK_SANCTIONED
- LENGTH_OF_ROAD_WORK_COMPLETED
- LENGTH_OF_ROAD_WORK_BALANCE
- NO_OF_ROAD_WORKS_SANCTIONED
- NO_OF_ROAD_WORKS_COMPLETED
- NO_OF_ROAD_WORKS_BALANCE
- NO_OF_BRIDGES_SANCTIONED
- NO_OF_BRIDGES_COMPLETED

- NO_OF_BRIDGES_BALANCE
- COST_OF_WORKS_SANCTIONED
- EXPENDITURE_OCCURED
- STATE_NAME, DISTRICT_NAME

Training Process:

- AutoAI split the dataset into training and holdout sets.
- Performed **preprocessing**, **feature engineering**, and **hyperparameter optimization** across multiple pipelines.
- 10 pipelines were generated; **Pipeline 10 (XGB + Ensemble)** ranked highest based on cross-validation.

Prediction Process:

- The trained model predicts the correct PMGSY scheme based on new project input.
- The model is deployed on IBM Cloud using **Watson Machine Learning**, enabling real-time predictions via API or UI.
- Supports integration into dashboards for practical government use.

RESULT

- Achieved a **highest accuracy of 92.4%** using the **Batched Tree Ensemble Classifier (XGBoost)** pipeline.

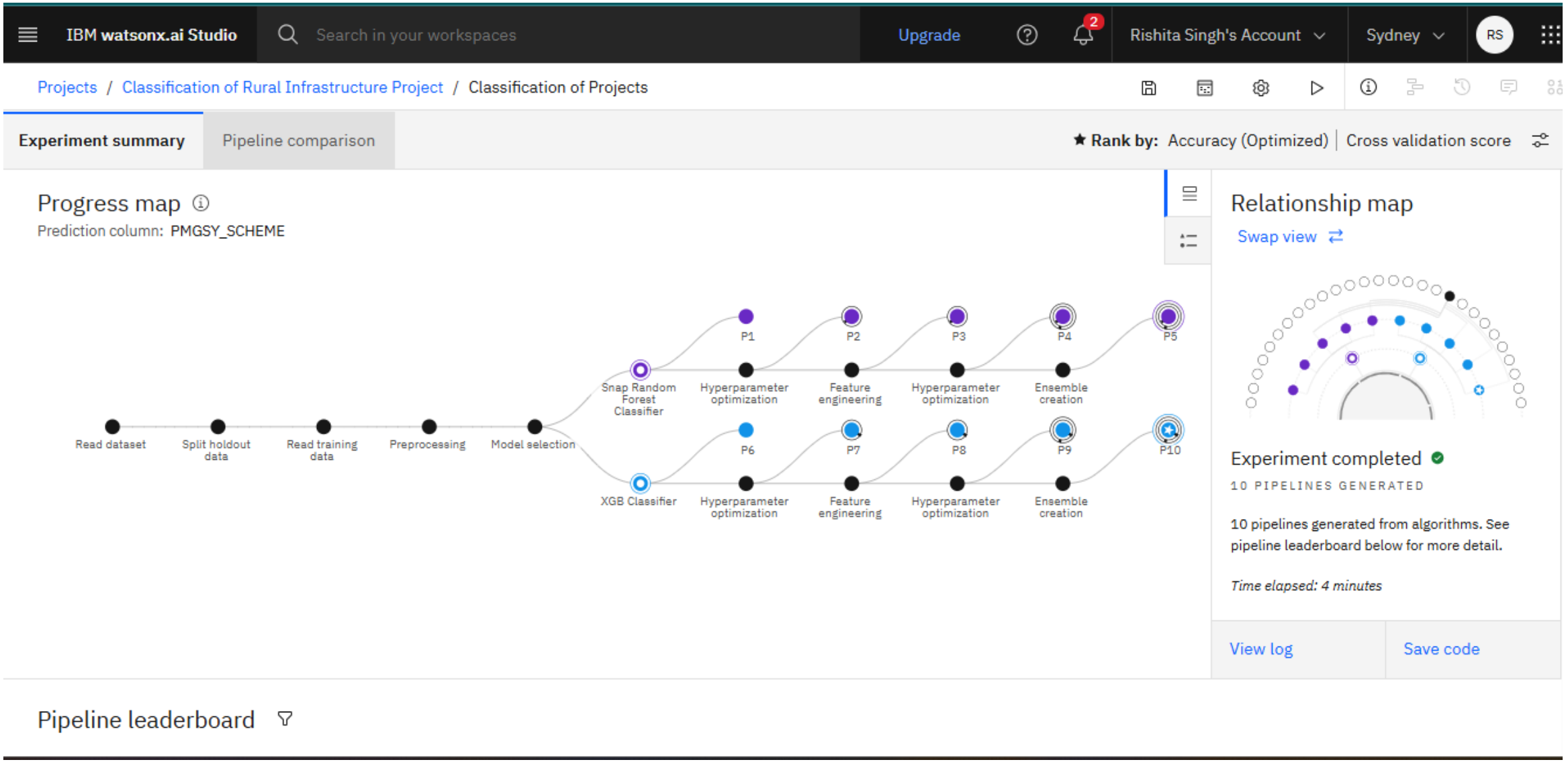
- Model was selected from **10 AutoAI-generated pipelines** based on cross-validation performance.

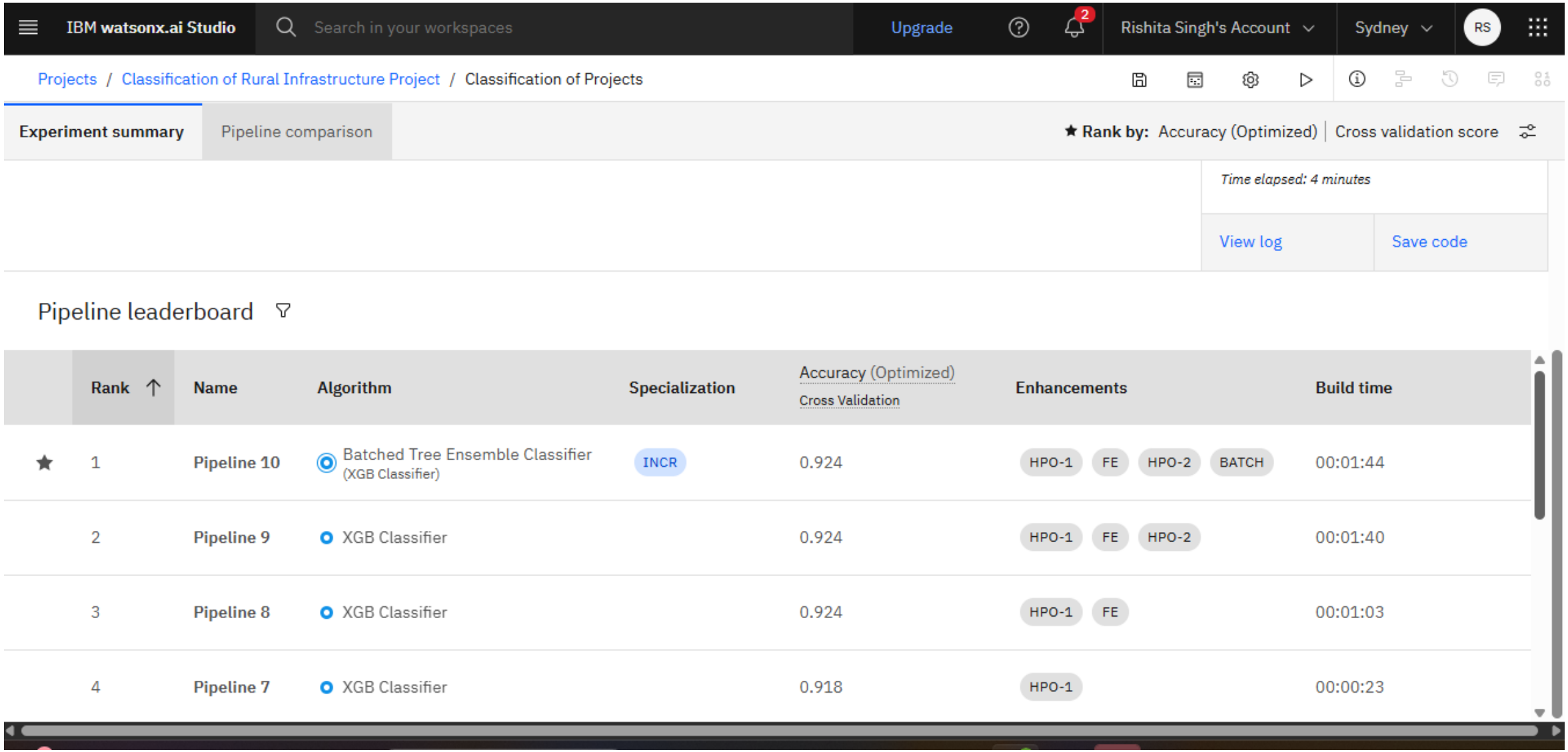
- Demonstrated strong accuracy in classifying projects into PMGSY-I, PMGSY-II, and RCPLWEA schemes.

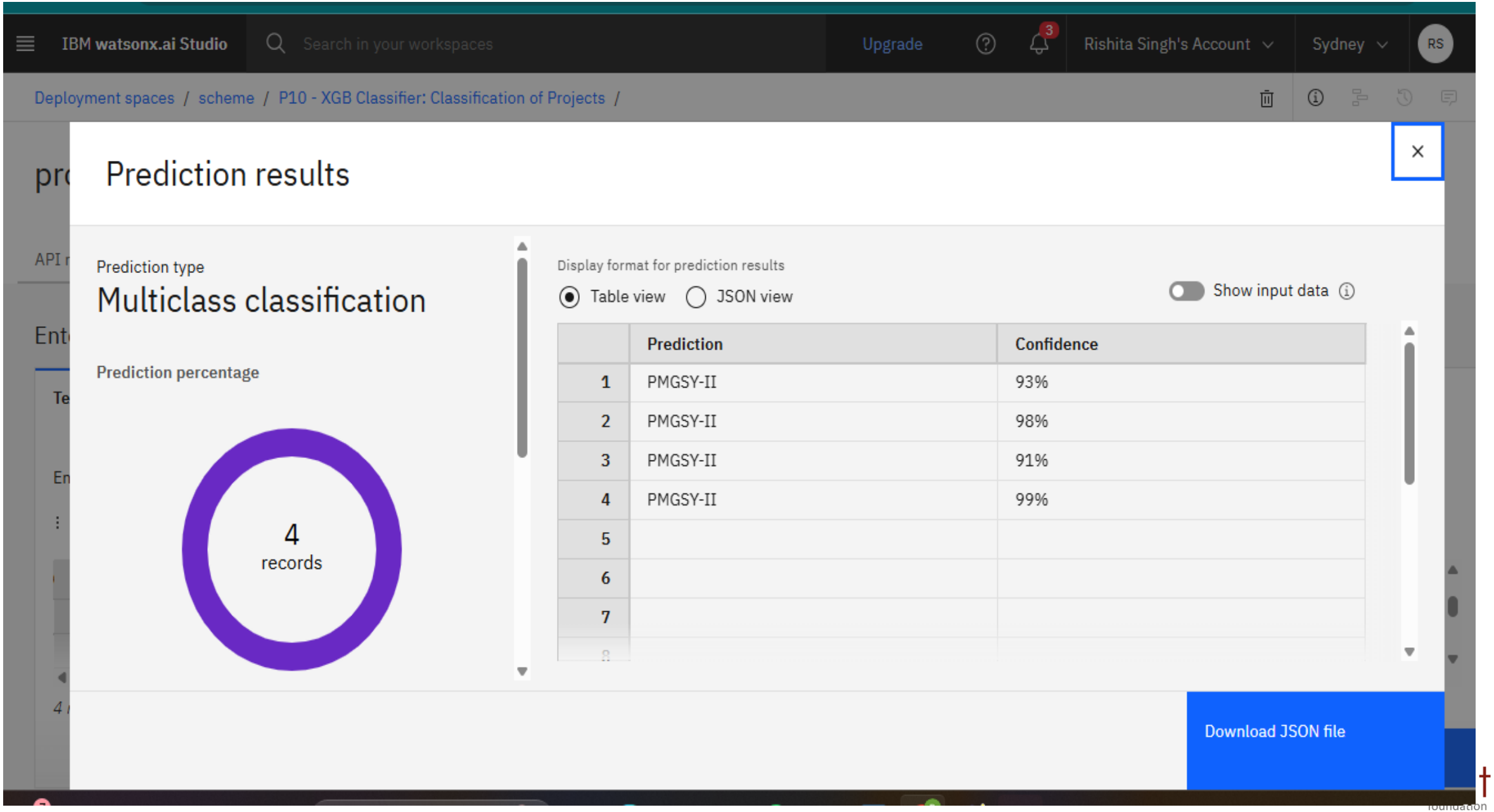
- Maintained high consistency across varied input features such as cost, length, and project completion status.

- Minor misclassifications occurred in overlapping scheme characteristics but within acceptable limits.

- Results confirm the model's effectiveness and readiness for real-world deployment.






IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Rishita Singh's Account
Sydney
RS
Deployment spaces / scheme / P10 - XGB Classifier: Classification of Projects /
Prediction results
Prediction type
Multiclass classification
Prediction percentage
4
records
Display format for prediction results
Table view
JSON view
Show input data
Prediction
Confidence
1 PMGSY-II 93%
2 PMGSY-II 98%
3 PMGSY-II 91%
4 PMGSY-II 99%
5
6
7
Download JSON file

CONCLUSION

- The machine learning model developed effectively classifies rural infrastructure projects into their respective PMGSY schemes based on key physical and financial features. This reduces manual workload, improves classification accuracy, and enhances transparency in monitoring and fund distribution.
- By deploying the solution on IBM Cloud, it ensures real-time accessibility, scalability, and easy integration with government systems.
- Key challenges such as data inconsistency and class imbalance were handled through preprocessing and model optimization.
- This project highlights the value of AI in public infrastructure planning and lays the groundwork for smarter, data-driven decision-making in rural development.

FUTURE SCOPE

- Expand the model with satellite imagery data for geospatial features
- Integrate with government PMGSY dashboards for live usage
- Incorporate NLP-based document classification from project reports
- Extend classification to new schemes or real-time project updates
- Use edge AI devices for rural field deployment
- Enable voice-based project data input for rural field officers using AI assistants.
- Integrate advanced AI models like transformers for document-based classification.

REFERENCES

- Ministry of Rural Development, PMGSY official portal
- Scikit-learn Documentation
- IBM Watson Studio Developer Resources
- Research paper: "Machine Learning Applications in Infrastructure Development", IEEE
- Data.gov.in – Government Open Data Platform

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Rishita 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

edunet foundation

THANK YOU

edunet foundation