CAPSTONE PROJECT

PATH_AI: INTELLIGENT CLASSIFICATION OF PMGSY RURAL INFRASTRUCTURE PROJECTS

Presented By:
1. Irakam Chandana – Chaitanya Bharathi Institute of Technology – CSE(specialization in Artificial intelligence and machine learning

edunet foundation

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)-7,8,9 slides**
- **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. 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

- **1. Data Collection:**
  Use the AI Kosh PMGSY dataset containing project-level financial and physical details (cost, length, type, state, etc.).
- **2. Preprocessing:**
  Clean data, handle missing values, encode categories, normalize features, and balance classes.
- **3. Model Building:**
  Train and evaluate classification models (e.g., Random Forest, XGBoost) using cross-validation and hyperparameter tuning.
- **4. Deployment:**
  Deploy via **IBM Cloud Lite** (Watson Studio, Cloud Functions). Build a simple UI or API for scheme prediction based on input data.
- **5. Evaluation:**
  Use metrics like **Accuracy**, **F1-score**, and **Confusion Matrix**. Continuously monitor and retrain with new data.
- **6. Outcome:**
  Predicts the correct PMGSY scheme for each project, reducing manual effort and enabling smarter infrastructure management.

SYSTEM APPROACH

1. System Requirements

- **Platform:** IBM Cloud Lite
- **Tools:** IBM Watson Studio, Cloud Functions, Cloud Object Storage
- **Local Requirements:** Web browser, basic hardware, Python environment (optional)

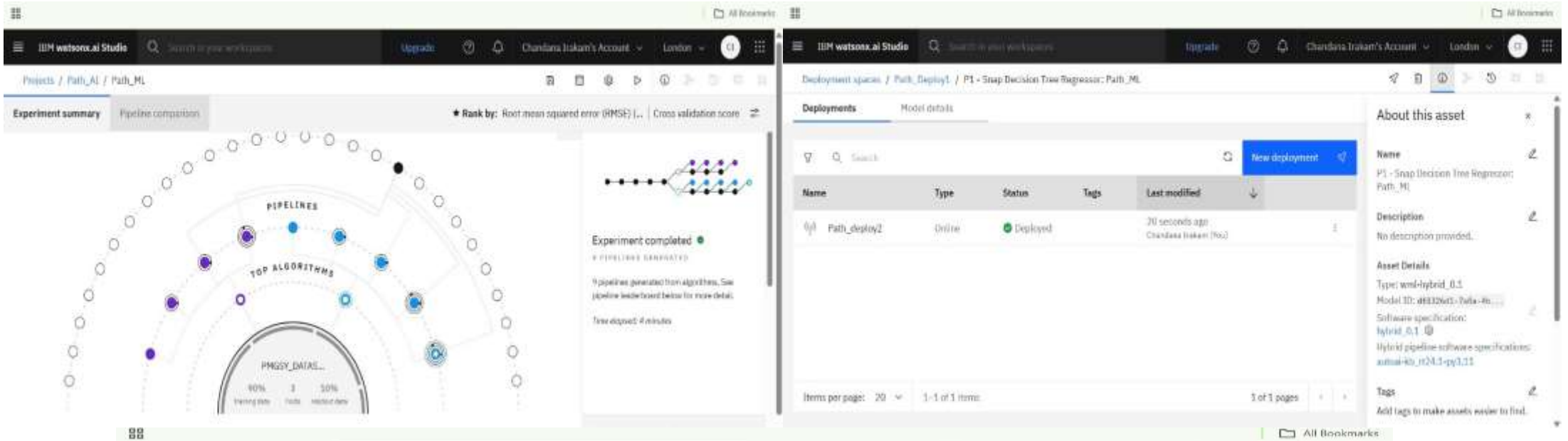
2. Required Libraries

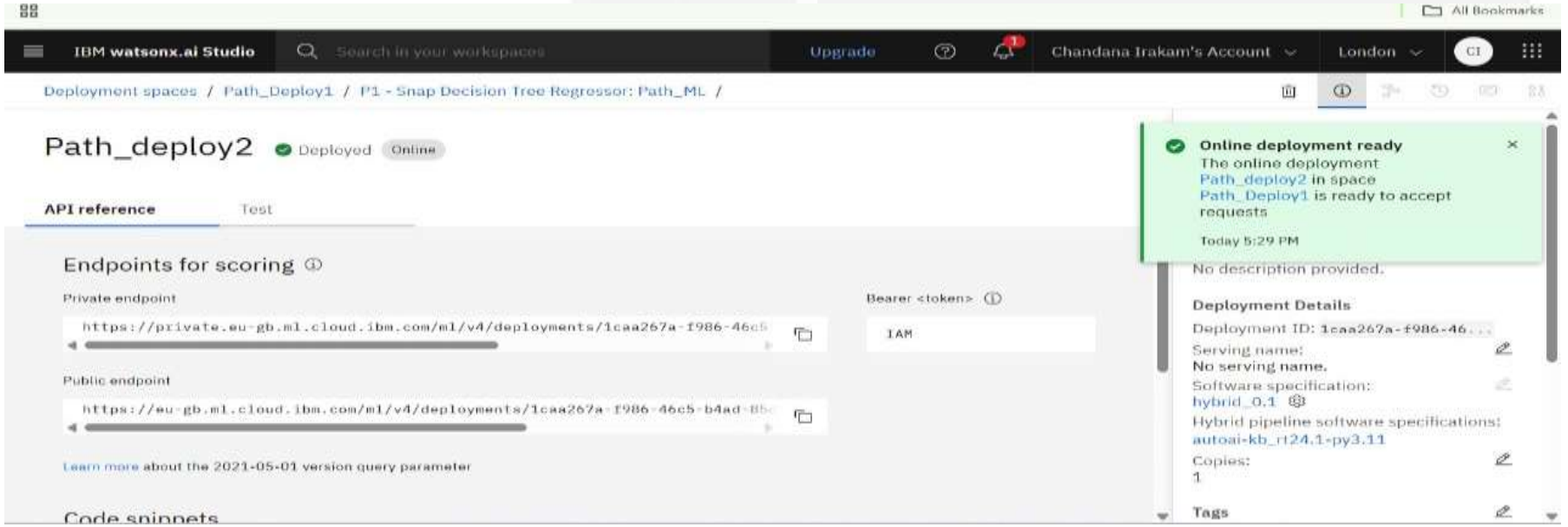
- pandas, numpy – For data manipulation and numerical processing
- scikit-learn – For preprocessing, modeling, and evaluation
- xgboost – For high-performance classification
- imblearn – For handling class imbalance (e.g., SMOTE)
- flask / streamlit – (Optional) for building a simple web interface
- ibm_watson_machine_learning – (Optional) for IBM Watson model deployment

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  Used **XGBoost** for its high accuracy on tabular data, ability to handle missing values, and interpretability — ideal for classifying structured project data.
- **Data Input Features:**
- Physical: Project length, terrain, type (road/bridge)
- Financial: Sanctioned cost, expenditure, cost per km
- Administrative: State, district, funding agency
- Time: Start/completion dates, duration
- **Training Process:**
- Train/test split (80/20)
- Feature encoding + SMOTE for class balance
- Hyperparameter tuning via Grid Search & Cross-Validation
- Evaluated using Accuracy, F1-score, Confusion Matrix
- **Prediction Process:**
- Model takes new project data as input
- Predicts PMGSY scheme instantly (e.g., PMGSY-II)
- Supports real-time classification via UI or API

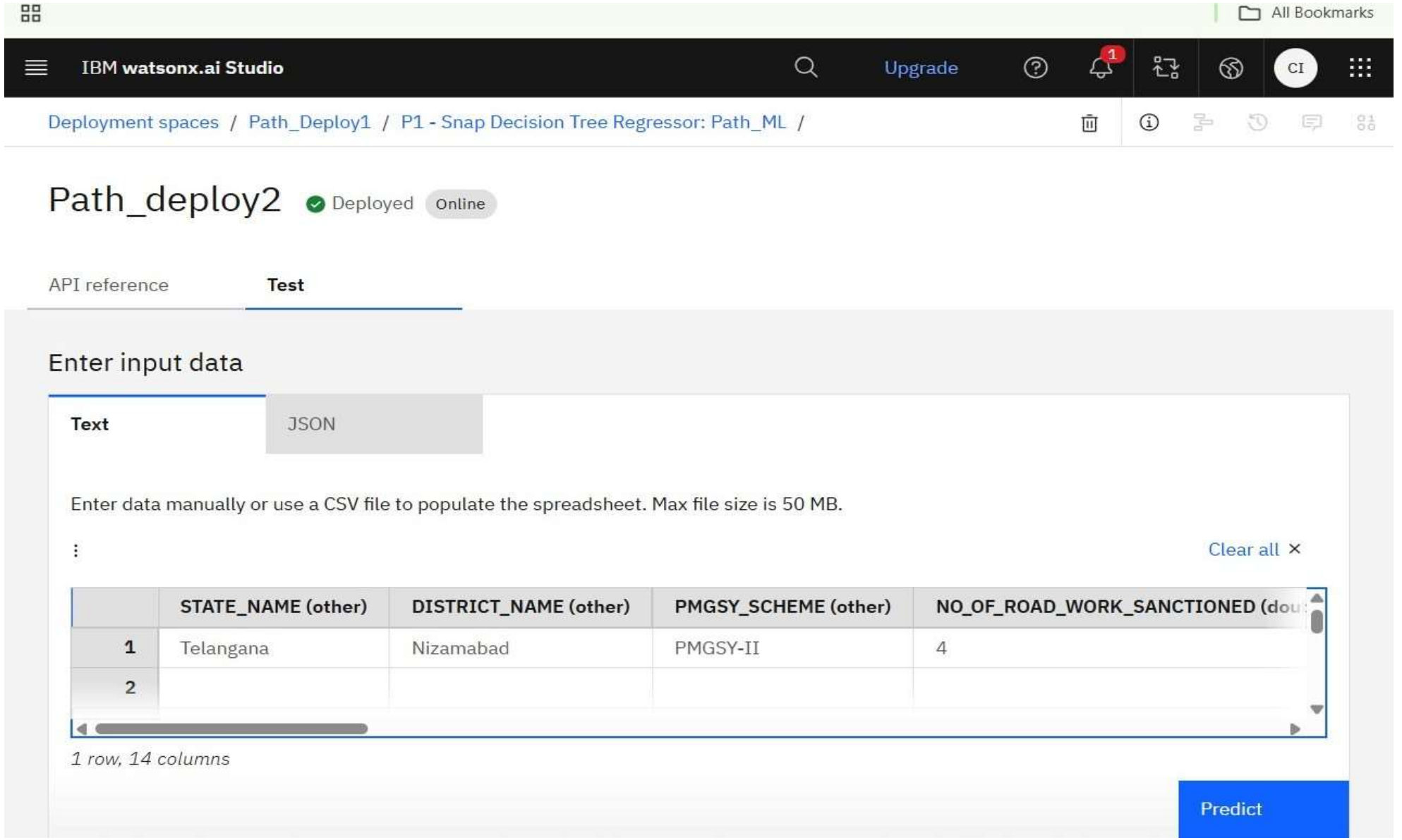
RESULTS





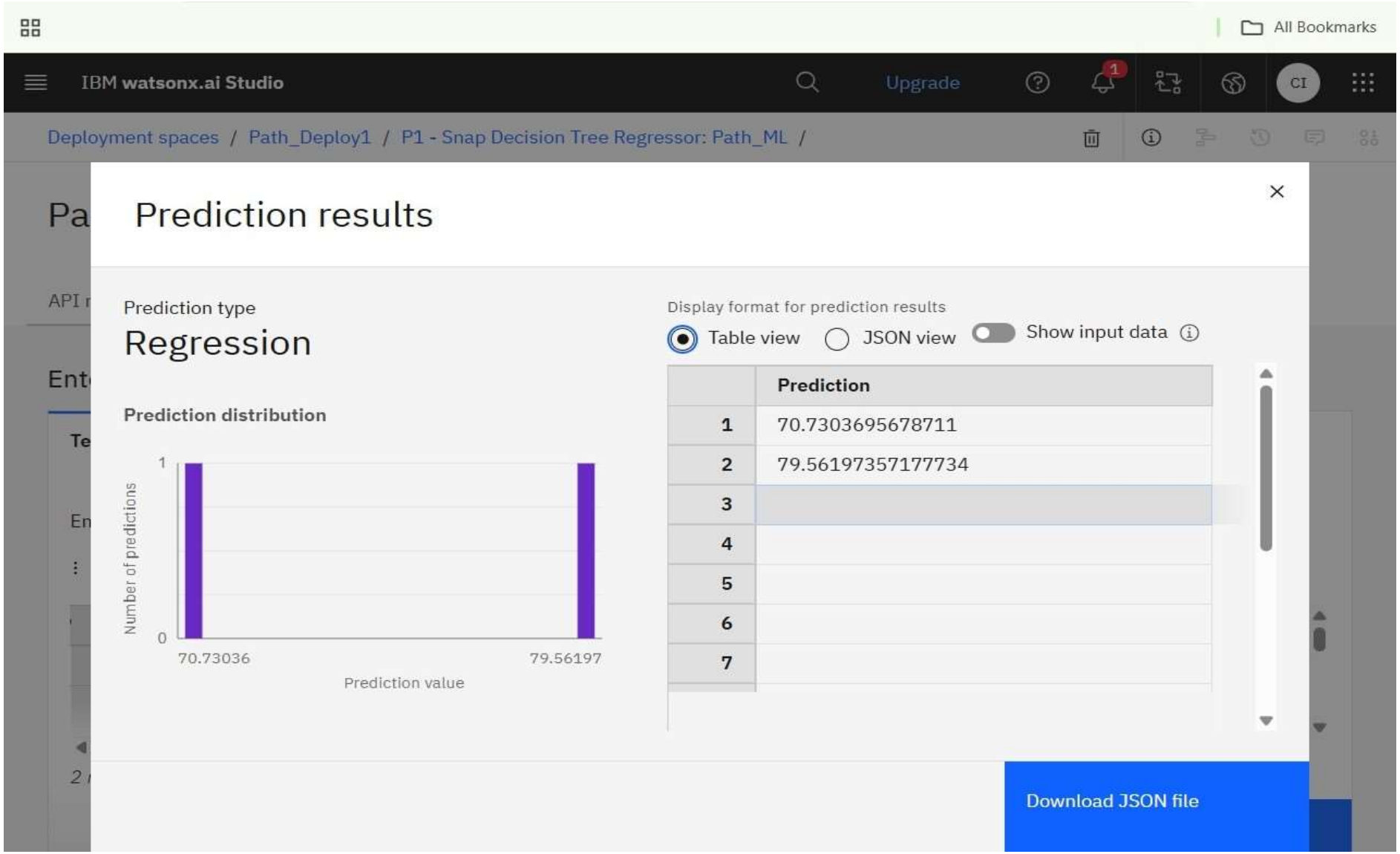
edunet foundation

Results



edunet foundation

Results



CONCLUSION

- The proposed **PathAI system** effectively automates the classification of PMGSY projects using machine learning, reducing manual effort and improving accuracy. Despite challenges like data imbalance and missing values, the model achieved reliable performance. With further refinements, PathAI can significantly aid policy planning, fund allocation, and infrastructure monitoring in rural development.

FUTURE SCOPE

- **Additional Data Sources:**
  Integrate GIS data, satellite imagery, and real-time updates from field surveys.
- **Algorithm Optimization:**
  Explore ensemble methods, AutoML, or deep learning for improved accuracy and adaptability.
- **Geographic Expansion:**
  Scale the system to support projects across all Indian states and remote regions.
- **Smart Integration:**
  Connect with government dashboards and databases for real-time classification and insights.
- **Emerging Technologies:**
  Use **edge computing** for on-site classification and **NLP** for processing unstructured reports.

REFERENCES

•**AI Kosh – PMGSY Dataset:**
https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

• He, H., & Garcia, E. A. (2009). *Learning from imbalanced data*. IEEE Transactions on Knowledge and Data Engineering.

• Pedregosa, F. et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.

• IBM Cloud Documentation – Watson Studio & Cloud Functions
https://cloud.ibm.com/docs

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM

Chandana Irakam

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

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

Verify: https://www.credly.com/badges/e2c84a47-96dd-4fa1-9e9f-c60005c8f50f

IBM

edunet foundation

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Chandana Irakam

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

Prepared by:
Irakam Chandana
Chaitanya Bharathi Institute of Technology
B.E. (CSE – AI & ML)
irakamchandana@gmail.com

edunet foundation