# SCHEME ELIGIBILITY PREDICTION FOR NSAP USING MACHINE LEARNING

*Presented By:*

**Niladri Chakraborty**
**Academy of Technology**
**MCA**

edunet foundation

# OUTLINE

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

# PROBLEM STATEMENT

The National Social Assistance Program (NSAP) is a flagship social security and welfare program by the Government of India. It aims to provide financial assistance to the elderly, widows, and persons with disabilities belonging to below-poverty-line (BPL) households. The program consists of several sub-schemes, each with specific eligibility criteria. Manually verifying applications and assigning the correct scheme is often time-consuming and error-prone. Such delays or incorrect allocations can prevent deserving individuals from receiving timely financial aid. This project focuses on designing and evaluating a multi-class classification model that accurately predicts the most appropriate NSAP scheme for an applicant based on their demographic and socio-economic data. The goal is to create a reliable tool that could assist government agencies in quickly and accurately categorizing applicants, ensuring that benefits are delivered to the right people efficiently.

# PROPOSED SOLUTION

The proposed system addresses the challenge of identifying the correct NSAP scheme for applicants by utilizing machine learning techniques on demographic and socio-economic data. The solution was implemented using IBM Cloud Watsonx.ai Studio, leveraging AutoAI for automated model building and optimization. The steps followed include:

- Data Collection & Preparation:
  - Imported NSAP scheme dataset containing demographic and socio-economic attributes, cleaned and preprocessed for analysis.
- Automated Model Building:
  - Used Watsonx.ai Studio's AutoAI to automatically generate multiple pipelines and evaluate various algorithms for multi-class classification of NSAP schemes.
- Model Selection & Optimization:
  - Snap Random Forest Classifier emerged as the best-performing model with an accuracy of 98.4% (Cross Validation) and 97.7% (Holdout).
- Evaluation & Deployment:
  - Model performance was validated using metrics like ROC curves, confusion matrix, and precision-recall curves. The finalized model was deployed on IBM Cloud as a REST API for real-time predictions.

This solution enables government agencies to automate scheme classification, reducing manual errors and ensuring timely financial support to eligible beneficiaries.

# SYSTEM APPROACH

To address the challenge of manual classification of NSAP schemes, a cloud-based automated machine learning workflow was implemented. The solution leverages IBM Cloud services to handle data processing, model building, evaluation, and deployment, ensuring scalability, automation, and easy integration with government systems.

- System Requirements:
  - IBM Cloud account (Lite plan) with Watsonx.ai Studio enabled
  - AutoAI for automated pipeline generation and evaluation
  - Cloud Object Storage for dataset hosting
  - Deployment space for hosting and testing APIs
- Libraries & Tools Used:
  - IBM Watsonx.ai Studio's AutoAI (built-in algorithms like Random Forest, Decision Tree)
  - Python-based environment (for model metrics & visualization)
  - REST API integration for model deployment

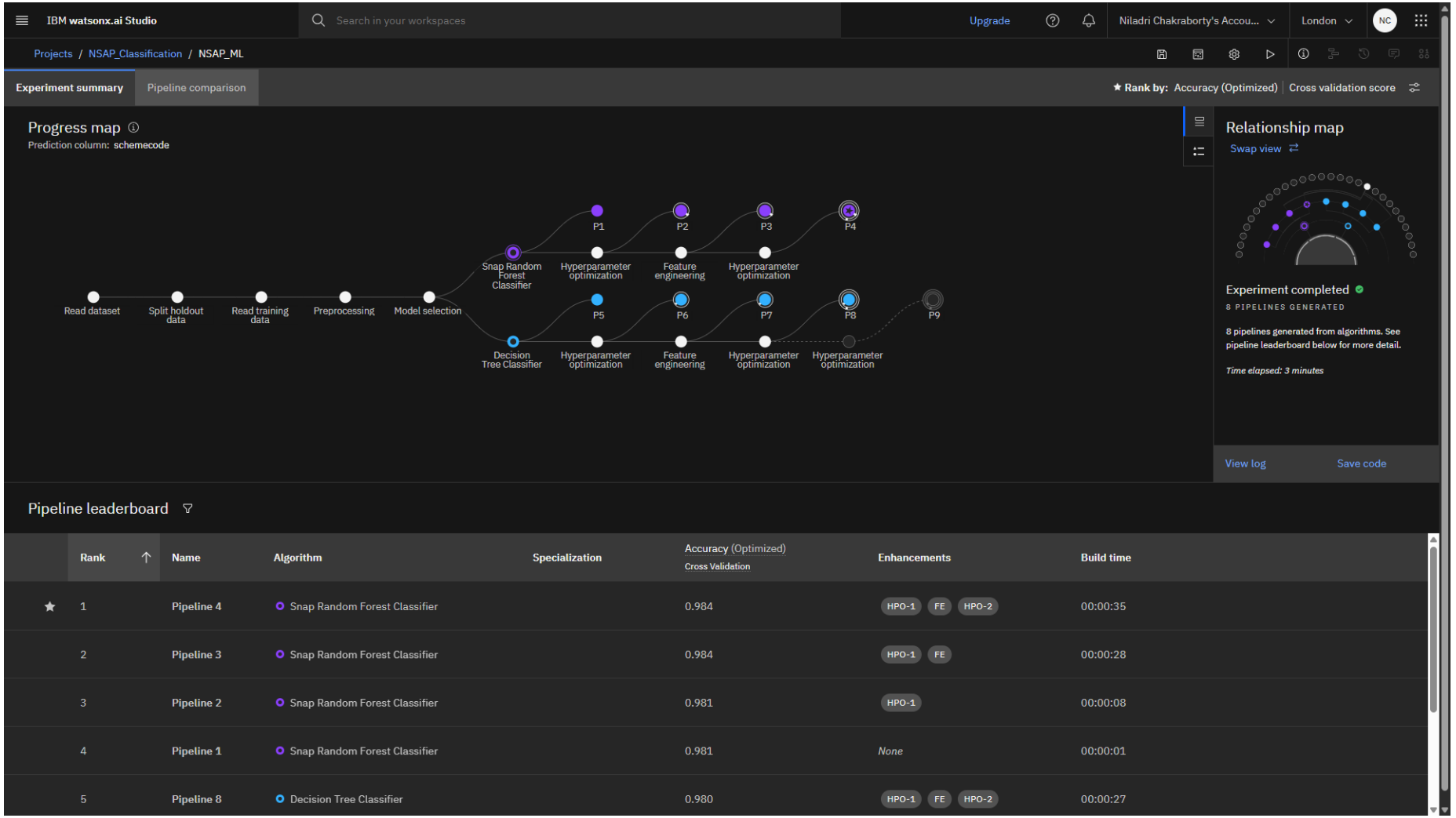
This approach enables the development of a robust, automated, and cloud-ready solution for efficient scheme classification and faster beneficiary support.

# ALGORITHM & DEPLOYMENT

- Algorithm Selection:
  - The Snap Random Forest Classifier generated by IBM AutoAI was chosen as the best-performing algorithm. Random Forest is an ensemble learning technique that builds multiple decision trees and merges their results to improve prediction accuracy and reduce overfitting. It is well-suited for multi-class classification tasks like predicting the appropriate NSAP scheme based on applicant data.
- Data Input:
  - The model used demographic and socio-economic attributes of applicants such as statename, districtname etc.
- Training Process:
  - The dataset was split into training and testing sets using an automated approach within IBM Watson AutoAI. Multiple pipelines were generated and evaluated based on metrics such as cross-validation accuracy. Hyperparameter tuning was automatically performed to select the best pipeline, resulting in 98.4% cross-validation accuracy and 97.7% holdout accuracy.
- Prediction Process:
  - The finalized Random Forest model predicts the correct NSAP scheme for each applicant based on their input attributes. After training, the model was deployed as a REST API on IBM Cloud, enabling real-time prediction for new applicant data directly from external systems.

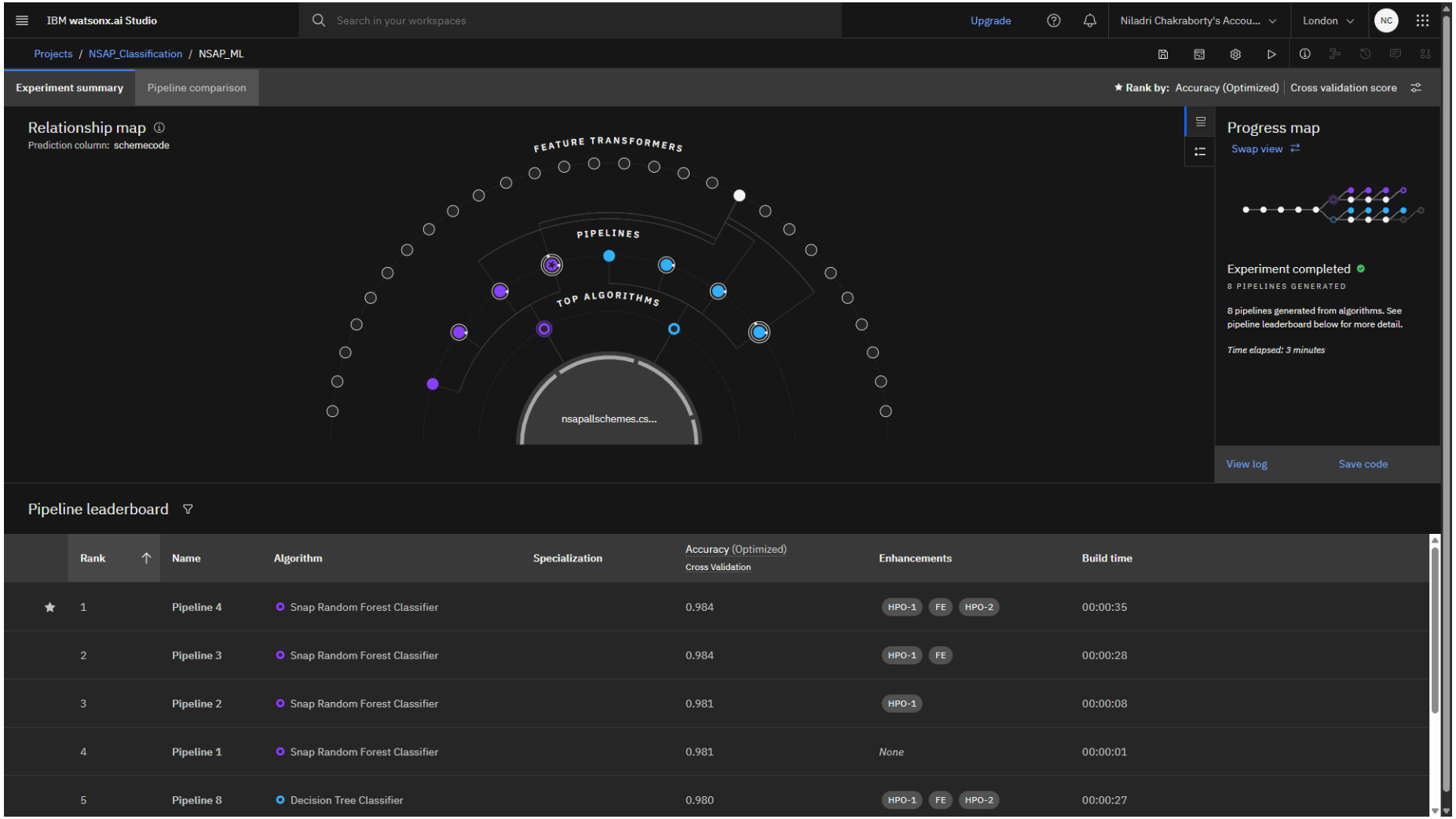
This deployment provides a highly accurate and automated classification solution, helping government agencies process applications faster and more reliably.

# RESULT




edunet foundation

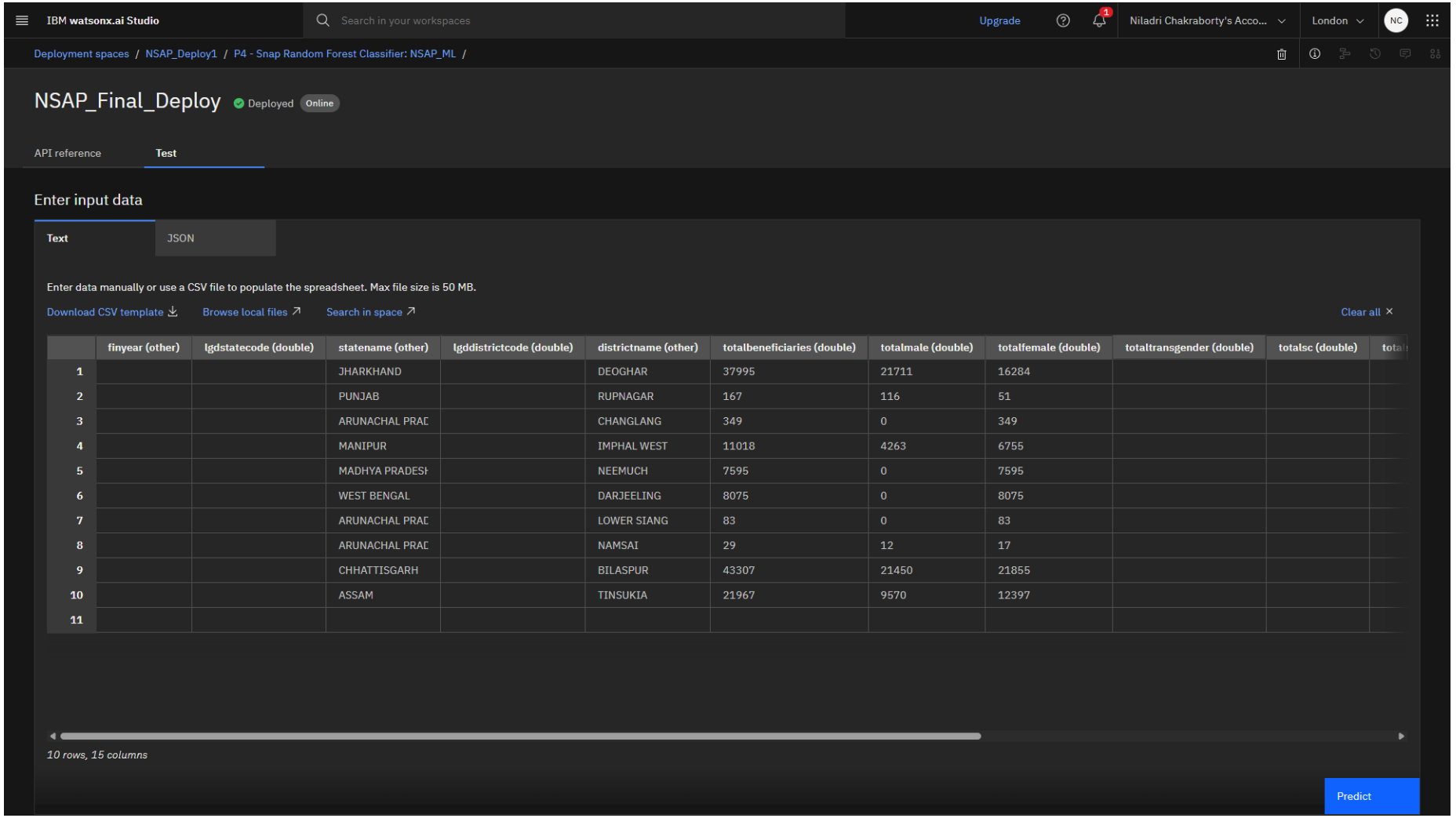
# RESULT




| | Rank | Name | Algorithm | Specialization | Accuracy (Optimized) Cross Validation | Enhancements | Build time |
|---|---|---|---|---|---|---|---|
| ★ | 1 | Pipeline 4 | Snap Random Forest Classifier | | 0.984 | HPO-1 FE HPO-2 | 00:00:35 |
| | 2 | Pipeline 3 | Snap Random Forest Classifier | | 0.984 | HPO-1 FE | 00:00:28 |
| | 3 | Pipeline 2 | Snap Random Forest Classifier | | 0.981 | HPO-1 | 00:00:08 |
| | 4 | Pipeline 1 | Snap Random Forest Classifier | | 0.981 | *None* | 00:00:01 |
| | 5 | Pipeline 8 | Decision Tree Classifier | | 0.980 | HPO-1 FE HPO-2 | 00:00:27 |

edunet foundation

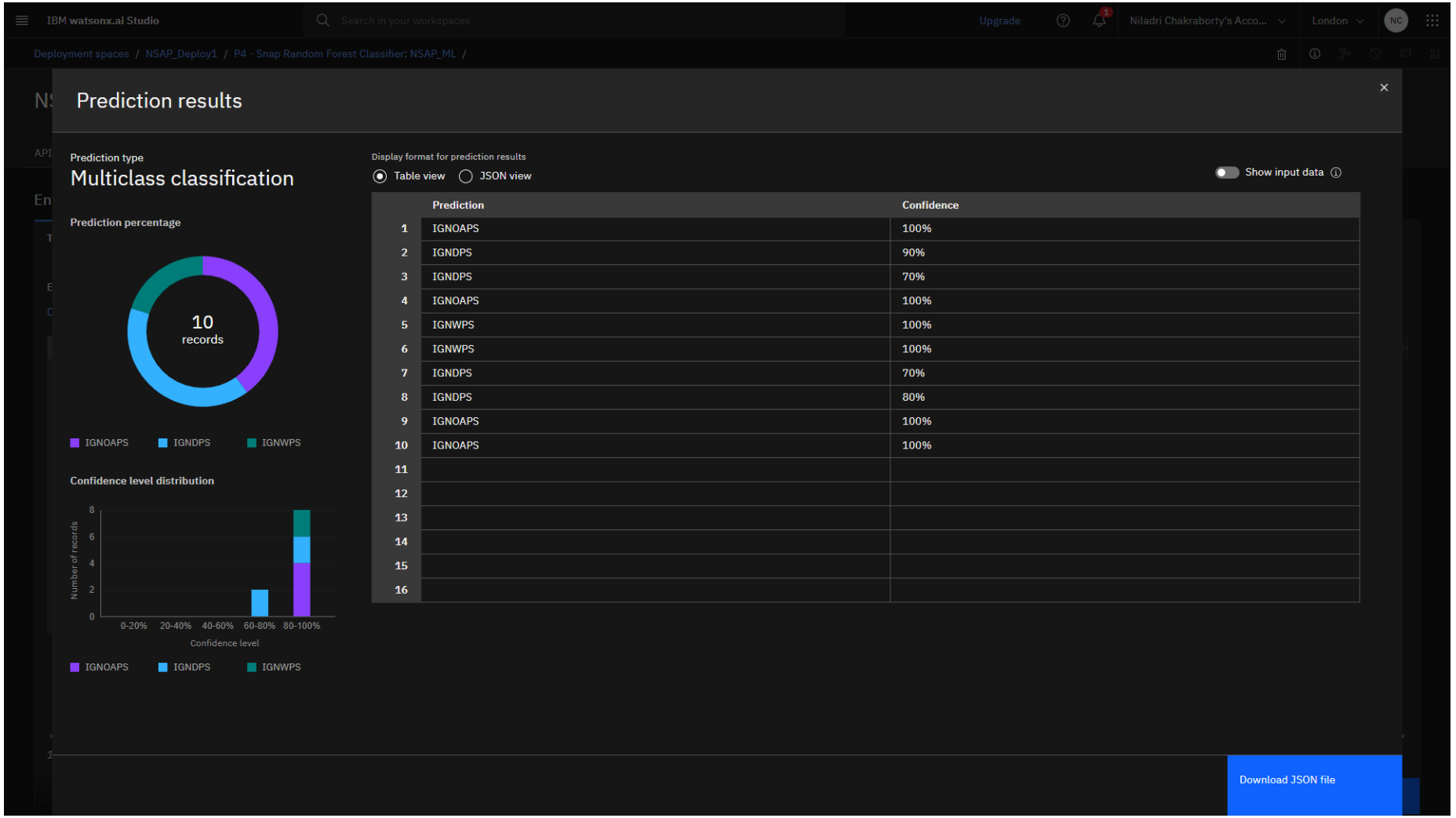
# RESULT




| | finyear (other) | lgdstatecode (double) | statename (other) | lgddistrictcode (double) | districtname (other) | totalbeneficiaries (double) | totalmale (double) | totalfemale (double) | totaltransgender (double) | totalsc (double) | tota |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | JHARKHAND | | DEOGHAR | 37995 | 21711 | 16284 | | | |
| 2 | | | PUNJAB | | RUPNAGAR | 167 | 116 | 51 | | | |
| 3 | | | ARUNACHAL PRAI | | CHANGLANG | 349 | 0 | 349 | | | |
| 4 | | | MANIPUR | | IMPHAL WEST | 11018 | 4263 | 6755 | | | |
| 5 | | | MADHYA PRADESH | | NEEMUCH | 7595 | 0 | 7595 | | | |
| 6 | | | WEST BENGAL | | DARJEELING | 8075 | 0 | 8075 | | | |
| 7 | | | ARUNACHAL PRAI | | LOWER SIANG | 83 | 0 | 83 | | | |
| 8 | | | ARUNACHAL PRAI | | NAMSAI | 29 | 12 | 17 | | | |
| 9 | | | CHHATTISGARH | | BILASPUR | 43307 | 21450 | 21855 | | | |
| 10 | | | ASSAM | | TINSUKIA | 21967 | 9570 | 12397 | | | |
| 11 | | | | | | | | | | | |

edunet foundation

# RESULT




| | Prediction | Confidence |
|---|---|---|
| 1 | IGNOAPS | 100% |
| 2 | IGNDPS | 90% |
| 3 | IGNDPS | 70% |
| 4 | IGNOAPS | 100% |
| 5 | IGNWPS | 100% |
| 6 | IGNWPS | 100% |
| 7 | IGNDPS | 70% |
| 8 | IGNDPS | 80% |
| 9 | IGNOAPS | 100% |
| 10 | IGNOAPS | 100% |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |

edunet foundation

# CONCLUSION

The project demonstrates the effectiveness of using cloud-based machine learning for automating the classification of NSAP schemes based on demographic and socio-economic applicant data. By leveraging IBM Cloud Watsonx.ai Studio and its AutoAI capabilities, a reliable model was developed and deployed as a REST API, enabling real-time predictions for scheme eligibility. This automation reduces manual intervention, minimizes human errors, and ensures timely delivery of financial support to deserving beneficiaries. Additionally, the cloud-based approach offers scalability, flexibility, and seamless integration with government systems, highlighting the potential of AI-driven solutions in improving public welfare services.

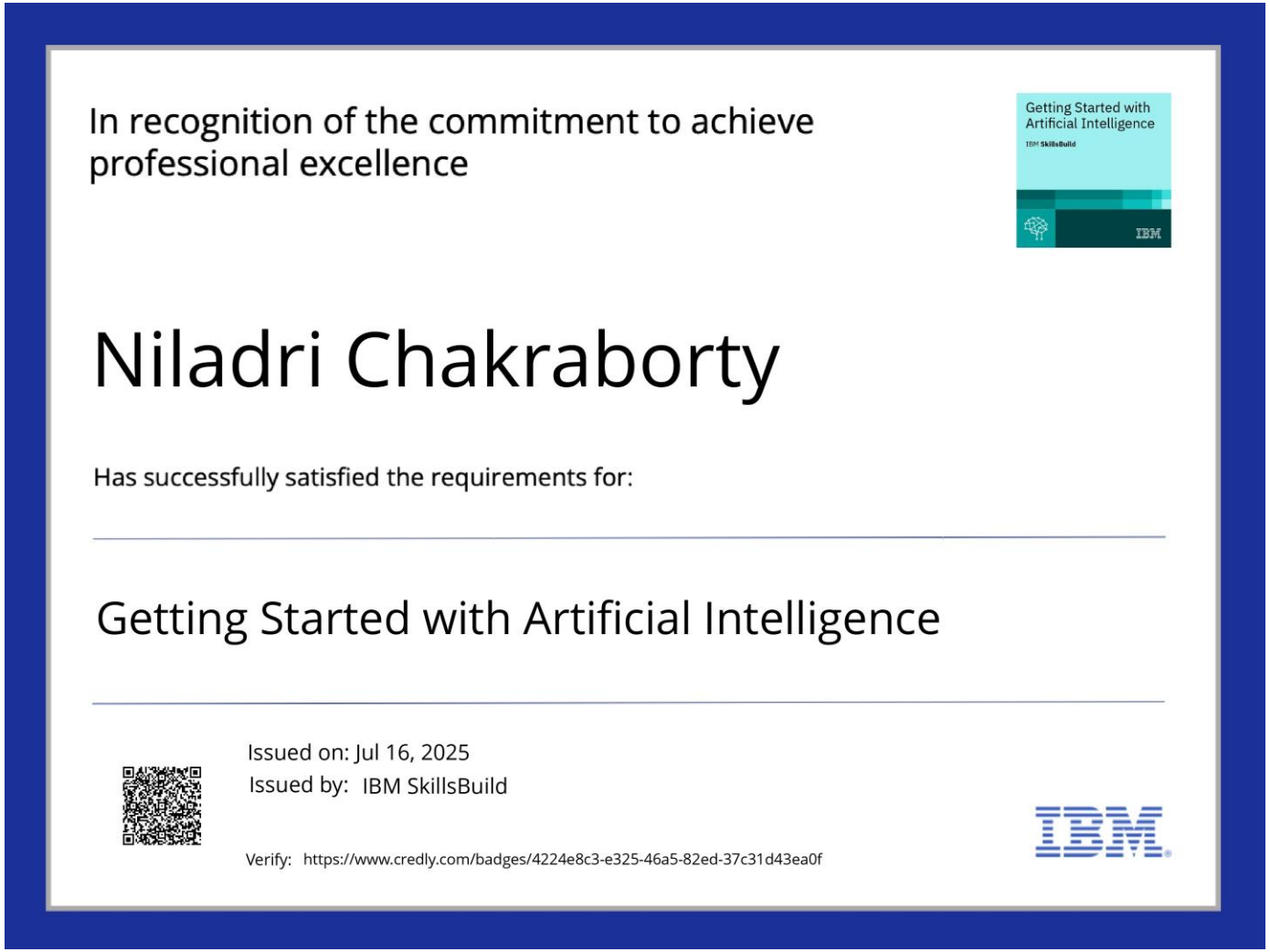
# FUTURE SCOPE

The system can be enhanced by incorporating additional and more diverse data sources, such as real-time socio-economic indicators and regional demographic updates, to further improve prediction accuracy. Advanced machine learning techniques and deep learning models can be explored to optimize performance and handle more complex eligibility criteria. The solution can also be scaled to cover multiple states and regions, enabling a nationwide deployment. Furthermore, integrating emerging technologies like edge computing for on-site data processing and advanced analytics for real-time decision-making can make the system even more efficient and adaptive to future government requirements.

# REFERENCES


- AI Kosh Dataset – District Wise Pension Data under the National Social Assistance Programme (NSAP)
- IBM Cloud Documentation – Watson Studio & AutoAI User Guide
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
- Government of India – National Social Assistance Programme Guidelines.

# IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Niladri Chakraborty

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/4224e8c3-e325-46a5-82ed-37c31d43ea0f

IBM

# IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

Niladri Chakraborty

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/bd01e0bb-33c9-494a-ab5d-f01862741fdf

IBM

# IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Niladri Chakraborty

for the completion of

**Lab: Retrieval Augmented Generation with LangChain**

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

**Completion date:** 22 Jul 2025 (GMT) **Learning hours:** 20 mins

# THANK YOU

edunet foundation