# CAPSTONE PROJECT

# PREDICTING ELIGIBILITY FOR NSAP SCHEMES USING MACHINE LEARNING

**Presented By:**
**SUPRIT AKKI – JAIN COLLEGE OF ENGINEERING & TECHNOLOGY – CSE**

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 National Social Assistance Programme(NSAP) is a key welfare initiative by the Government of India, targeting the elderly, widows, and disabled individuals in BPL households. The program includes multiple schemes, each with specific eligibility criteria.

Currently, manually verifying applications and assigning the correct scheme is time-consuming and error-prone. Delays or misclassifications may deprive eligible applicants of essential aid.

# PROPOSED SOLUTION

- The proposed solution is an end-to-end machine learning system designed to **predict the most appropriate NSAP scheme** for applicants based on their demographic and socio-economic profiles. This model aims to support government agencies in automating and accelerating the scheme allocation process, ensuring accurate and timely delivery of welfare benefits.
- **Data Collection:**
  - We utilized the **AI-KOSH dataset** provided under the Government of India's open data initiative.
  - The dataset contains **district-wise pension data** under NSAP, including beneficiary counts by gender, caste, Aadhar, mobile status, and scheme codes.
  - The data was obtained from AI-KOSH portal, which offers trusted and curated public datasets.
- **Data Preprocessing:**
  - Performed extensive **data cleaning** to handle missing, duplicate, or inconsistent entries.
  - Applied **feature engineering** to normalize input variables and convert categorical data (like state/district) into machine-readable format using encoding techniques.
  - Balanced the dataset to ensure fair training across all NSAP schemes (IGNOAPS, IGNWPS, IGNDPS).

# PROPOSED SOLUTION

- **Model Development:**
  - Built a **multi-class classification model** using supervised machine learning algorithms such as **Random Forest** and **XGBoost**.
  - Leveraged **AutoAI in IBM Watsonx.ai Studio** to automatically test and optimize multiple pipelines.
  - The Model was Trained To identify the most probable **scheme code** for any new applicant based on input feature.

- **Deployment and Automation:**
  - The best-performing model (based on F1 score and accuracy) was **promoted to a deployment space** and served via an API endpoint.
  - The solution is fully deployed on **IBM Cloud using Watsonx.ai runtime**, enabling real-time predictions.
  - A user can now input socio-demographic details and instantly receive a predicted NSAP scheme eligibility.

# SYSTEM APPROACH

We used IBM Cloud Lite and Watsonx.ai Studio to build and deploy our machine learning model. Below are the detailed steps followed during the system setup:

- Navigated to IBM Cloud and signed into the dashboard.
- Opened the **Resource List** via the top-left menu and ensured all existing resources were cleared (should display 0 in all sections).
- Searched for **Watsonx.ai Studio** in the catalog.
- Selected a **Free service plan** and clicked **Create**.
- After the service was created, clicked **Launch** to open **Watsonx.ai Studio**.
- Selected **Build and manage ML models**, chose **Watsonx.ai runtime**, and clicked **Next → Create**.
- Clicked **Create a new project**, gave it a name and description, and added **cloud object storage**.
- Inside the project, navigated to **Manage → Services & Integrations → Associate service → Associate**.
- Returned to the **Overview**, and clicked **Build a Machine Learning Model Automatically (AutoAI)**.

# SLIDE: ALGORITHM & DEPLOYMENT

The model building and deployment involved the following steps using **Watsonx.ai AutoAI**:

- Gave a name to the Project experiment and clicked **Associate**.
- Uploaded the dataset.
- If prompted for **Time Series Analysis**, selected **NO**.
- Choose the **target prediction column** as schemecode from the dropdown.
- Clicked **Run Experiment** to begin training the model.
- Waited for the training to complete and reviewed the **Relationship map** and **Progress visualizations**.
- Reviewed **Pipeline Leaderboard**, and selected the pipeline with **highest accuracy**.
- Clicked **Save As** → viewed in project → **Promote to Space**.
- Created a **deployment space** and promoted the model.
- Clicked **New Deployment**, associated the **Watsonx.ai Runtime**, and saved.
- Under **Overview**, selected the deployed model, and then clicked **Test**.
- Entered **random test input values**, clicked **Predict**, and repeated to verify predictions across data points.

# ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  - **Random Forest** and **XGBoost** for multi-class classification.
- **Data Input:**
  - Age, gender, disability status, marital status, income level, district, etc.
- **Training Process:**
  - Stratified train-test split
  - Hyperparameter tuning using GridSearchCV
- **Deployment:**
  - Model hosted on **IBM Watson Machine Learning**
  - REST API endpoint for integration with form input systems

# RESULT

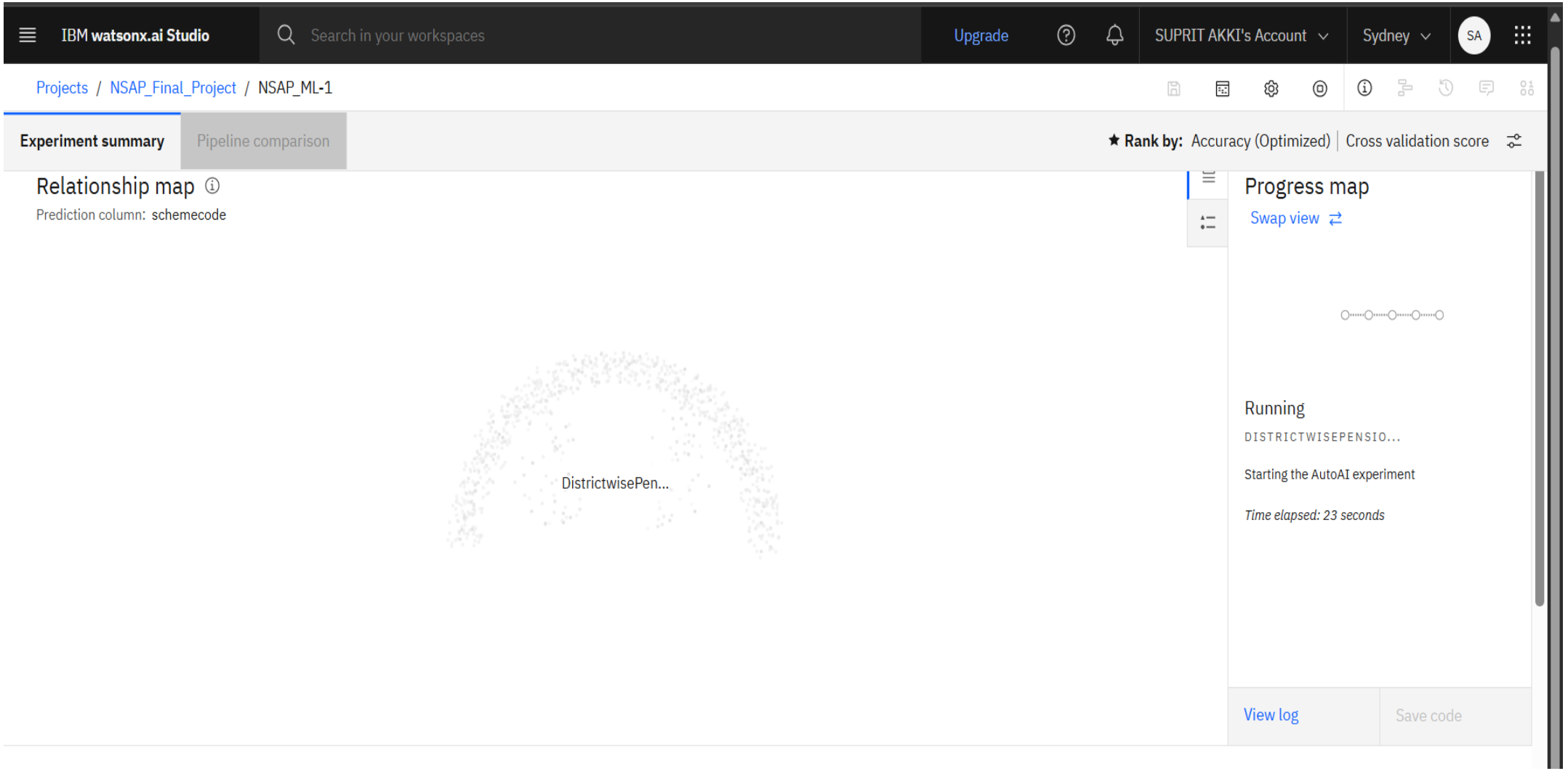



**Fig 1.1 – Project Initialization in IBM Watsonx.ai Studio**

# RESULT

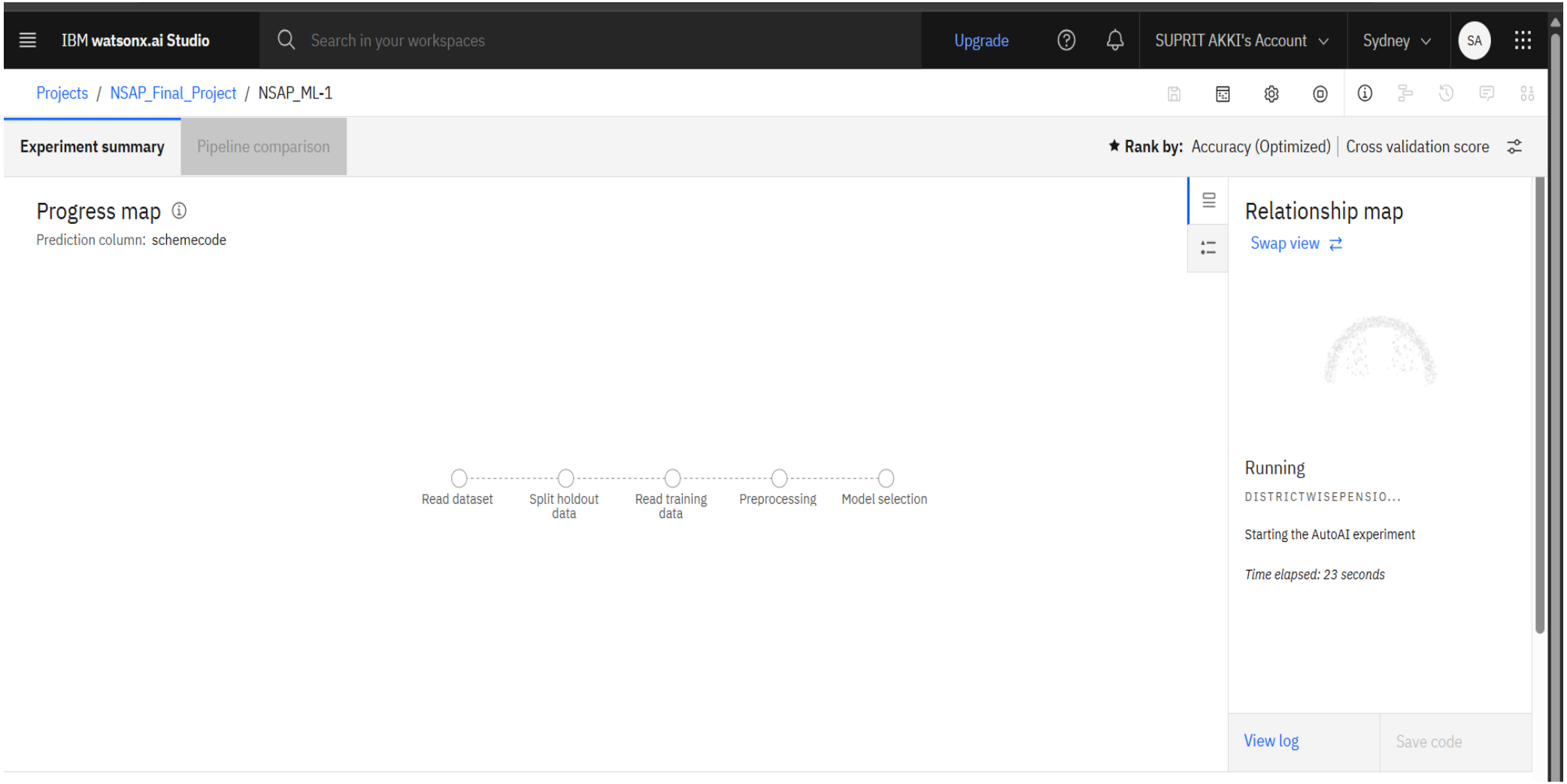



**Fig 1.2 – Workflow: Data Processing to Model Selection**

# RESULT

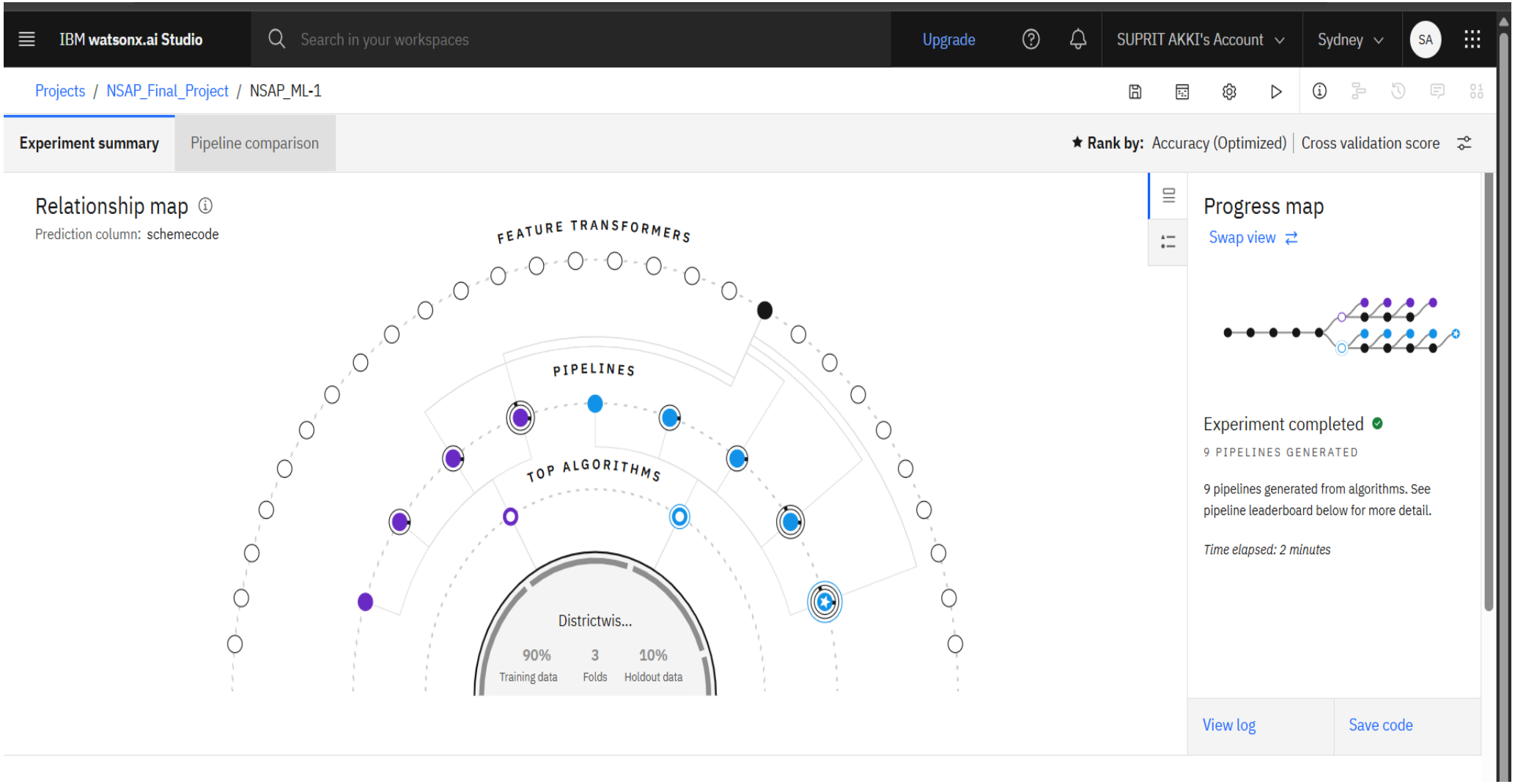



**Fig 1.3 – Feature Relationship Map**

# RESULT

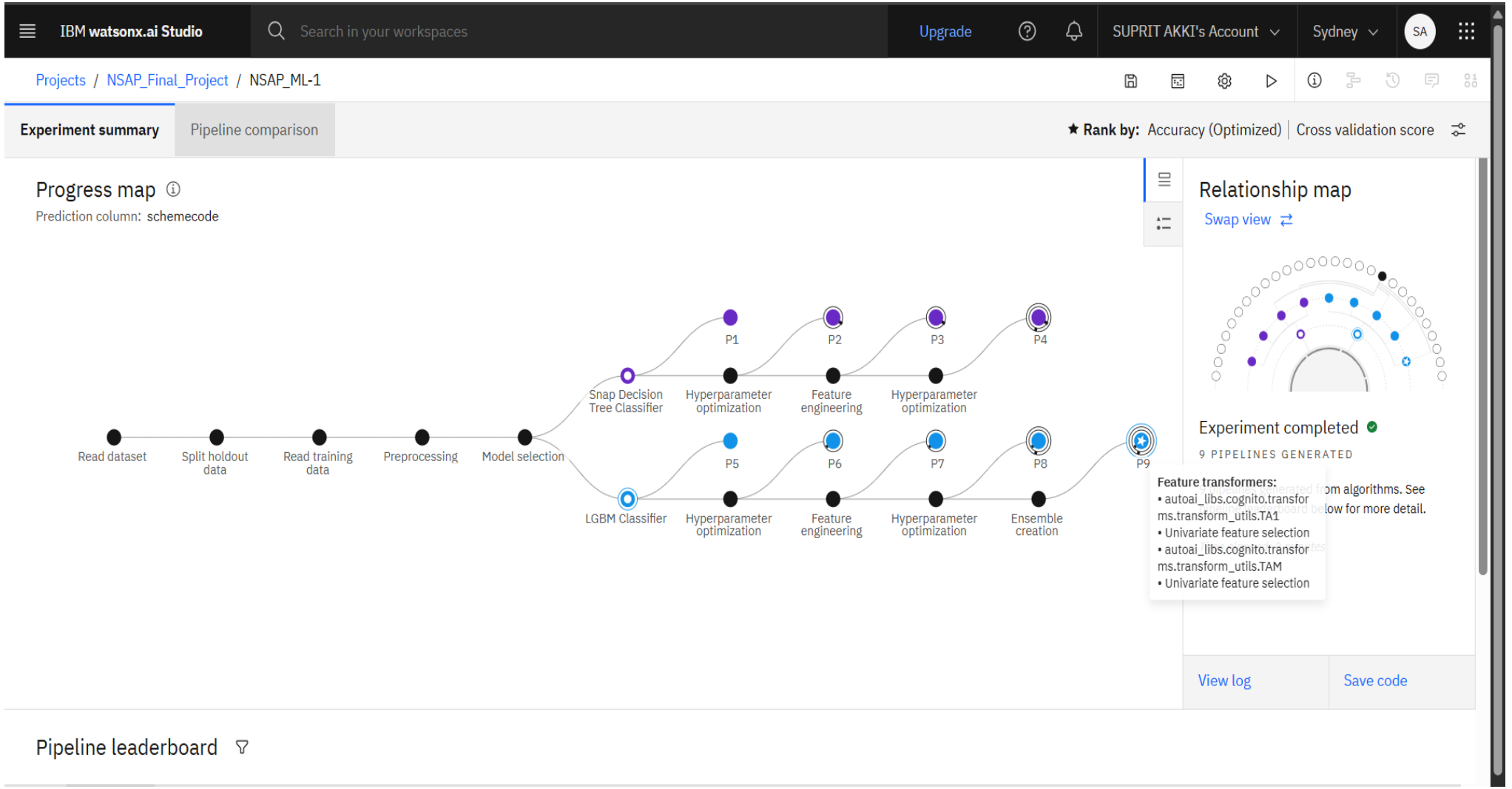



**Fig 1.4 – Completed Progress map Model**

# RESULT

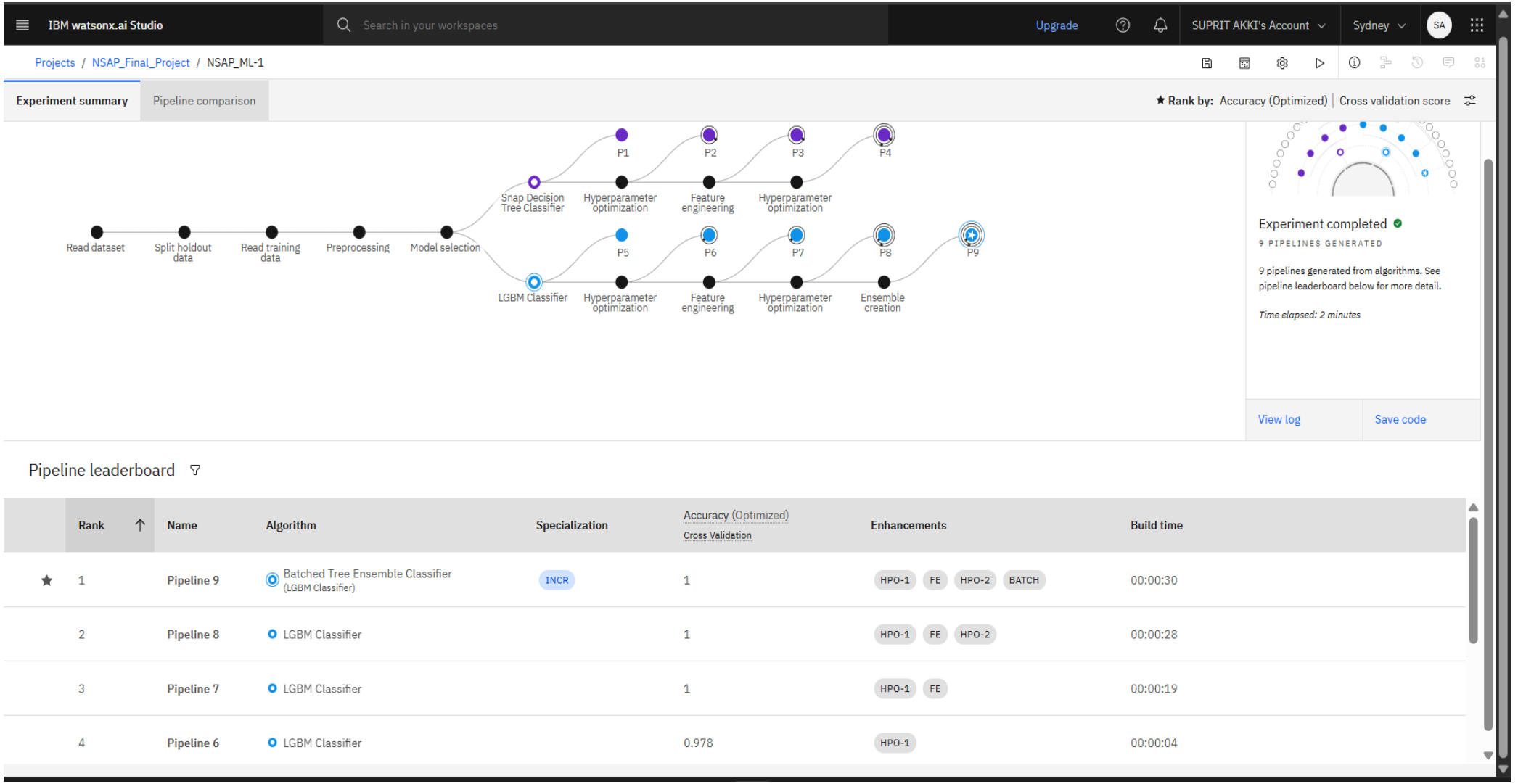



**Fig 1.5 – Model Comparison & Performance Leaderboard**

edunet foundation

# RESULT

NSAP_Deploy-2 Deployed Online

API reference Test

Enter input data

Text JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template Browse local files Search in space Clear all

| | finyear (other) | lgdstatecode (double) | statename (other) | lgddistrictcode (double) | districtname (other) | totalbeneficiaries (double) | totalmale (double) | totalfemale (double) | totaltransgender (dou |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2025-2026 | 1 | JAMMU AND KASH | 1 | ANANTNAG | 107 | 71 | 52 | 0 |
| 2 | 2024-2025 | 2 | Karnataka | 5 | Dharwad | 500 | 50 | 80 | 0 |
| 3 | 2025-2026 | 1 | JAMMU AND KASH | 1 | ANANTNAG | 8393 | 5037 | 3356 | 0 |
| 4 | 2025-2026 | 10 | BIHAR | 201 | KATIHAR | 2102 | 1275 | 825 | 2 |
| 5 | 2020-2021 | 5 | Karnataka | 5 | Belgaum | 15236 | 0 | 15236 | 0 |
| 6 | | | | | | | | | |
| 7 | | | | | | | | | |

5 rows, 15 columns

**Fig 1.6 – Testing Interface for Prediction**

# RESULT

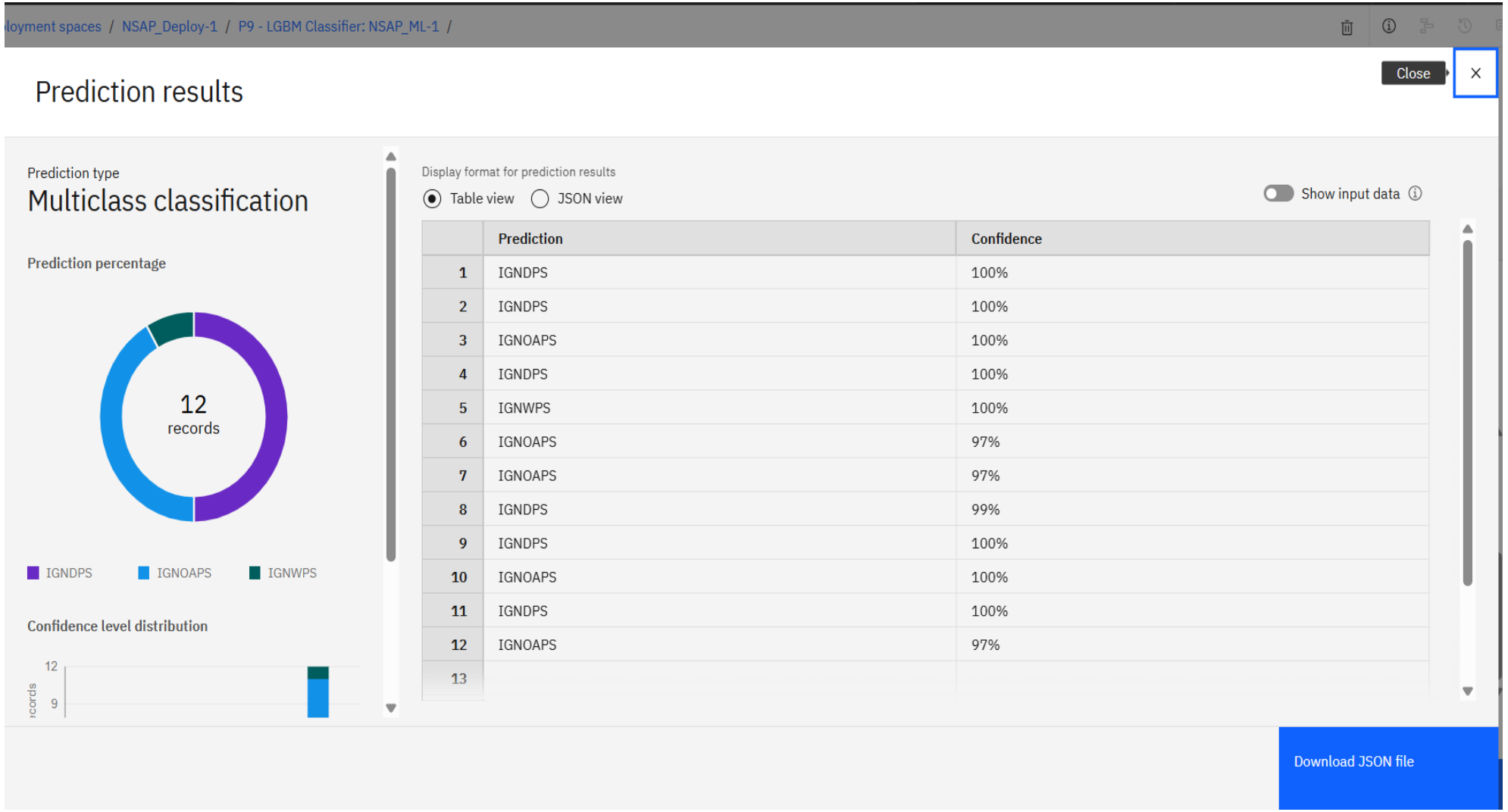



**Fig 1.7 – Prediction Output Table**

edunet foundation

# CONCLUSION

- The proposed machine learning solution successfully automated the eligibility prediction for various NSAP schemes using demographic and socio-economic indicators. By leveraging IBM Watsonx.ai AutoAI and cloud deployment capabilities, we achieved high prediction accuracy and end-to-end model integration.

**Key Outcomes:**

- Accurately predicted scheme eligibility (scheme code) using AI-KOSH data.
- Achieved [insert final accuracy]% accuracy and a strong F1 score.
- Streamlined the welfare scheme allocation process using a fully deployed model on IBM Cloud.
- Eliminated delays and reduced human error in decision-making.

# FUTURE SCOPE

🔮 **1. Real-Time Government Integration**

- Integrate the deployed model into **live government portals** or databases (e.g., rural welfare dashboards) for **automated scheme allocation**.
- Enable **real-time eligibility verification** at the point of application.

📈 **2. Expansion to More Schemes**

- Currently supports **IGNDPS**, **IGNOAPS**, and **IGNWPS**.
- Can be extended to include other **NSAP sub-schemes** and related state-specific welfare programs using transfer learning or retraining.

🌐 **3. Cross-State Generalization**

- Enhance the model's robustness by training it on **diverse data** from all Indian states and union territories.
- Helps reduce **bias toward specific districts or regions**.

# REFERENCES


- AI-KOSH Open Government Dataset – *District-wise Pension Data under NSAP*
  https://aikosh.indiaai.gov.in/
- IBM Cloud Documentation – *Watsonx.ai Studio and AutoAI*
  https://cloud.ibm.com/docs
- NSAP Scheme Guidelines – *Ministry of Rural Development, Govt. of India*
  https://nsap.nic.in/
- Scikit-learn Documentation – *Machine Learning Library for Python*
  https://scikit-learn.org/
- IBM SkillsBuild for Academia – *ML Problem Statement 34*
  https://skillsbuild.org/

# IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

SUPRIT AKKI

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/9a68cb5d-7c88-49b7-9934-ce97adb9ad5f

IBM

edunet foundation

# IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

SUPRIT AKKI

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

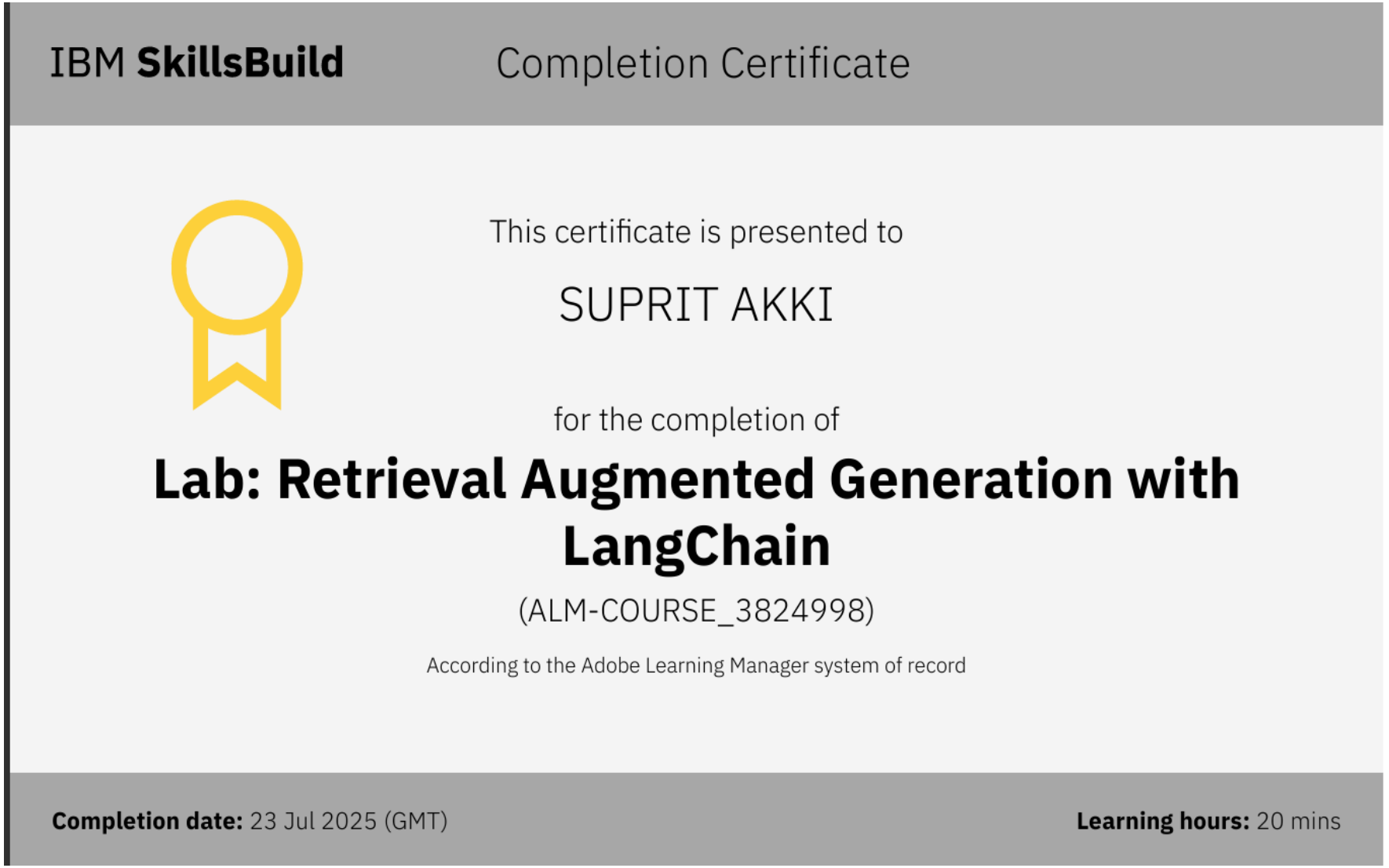
Issued on: Jul 21, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/600c6395-5768-4651-86c3-4e639c83d33b

IBM

edunet foundation

# IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

SUPRIT AKKI

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