CAPSTONE PROJECT

PREDICTING NSAP SCHEME ELIGIBILITY USING MACHINE LEARNING

Presented By:
KudalaSrigurunath
Sreyas Institute of Engineering and Technology
Department of Computer Science and Engineering (Data Science)

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

The **National Social Assistance Program (NSAP)** is a vital welfare initiative by the Government of India aimed at providing financial support to the **elderly**, **widows**, and **persons with disabilities** belonging to **below-poverty-line (BPL)**households. The program consists of multiple sub-schemes, each governed by specific eligibility criteria tailored to different beneficiary groups.

Currently, the process of manually verifying applications and assigning the correct scheme is **labour-intensive**, **error-prone**, and **time-consuming.** These inefficiencies can lead to **delays**, **misclassification**, or even **denial of aid** to deserving individuals. Such issues not only hinder the efficiency of the welfare system but also negatively impact **social equity**, **trust**, and **timely delivery** of benefits

PROPOSED SOLUTION

- The proposed system aims to automate the eligibility classification process for NSAP schemes using a no-code machine learning approach. By leveraging IBM Watsonx.ai AutoAI, the system can accurately classify scheme eligibility based on district-level demographic and socio-economic data. The solution consists of the following components:
- Data Collection:

  Gather district-wise NSAP data from the AI Kosh portal, including gender, caste, Aadhaar linkage, and mobile linkage.

  Use scheme code as the target column representing the NSAP scheme to be predicted.
- Data Preprocessing:

  Auto AI automatically detects data types, handles missing values, and encodes categorical variables.

  Important features like total male, total female, totalsc,totalst,totalbc,totalgen(caste) , total Aadhaar, and total mobile number are extracted and transformed
- Machine Learning Algorithm:

  Auto AI builds multiple models using algorithms such as Snap Random Forest, Logistic Regression, and Gradient Boosting.

  It applies hyperparameter tuning (HPO-1, HPO-2) and ranks models based on evaluation metrics like F1-score and accuracy
- Evaluation:

  Auto AI generates a leaderboard comparing model performance using metrics like accuracy, precision, recall, and F1-score.

  The final model (Snap Random Forest) achieved a high accuracy of **98.4%**, ensuring reliable scheme classification.
- Result:

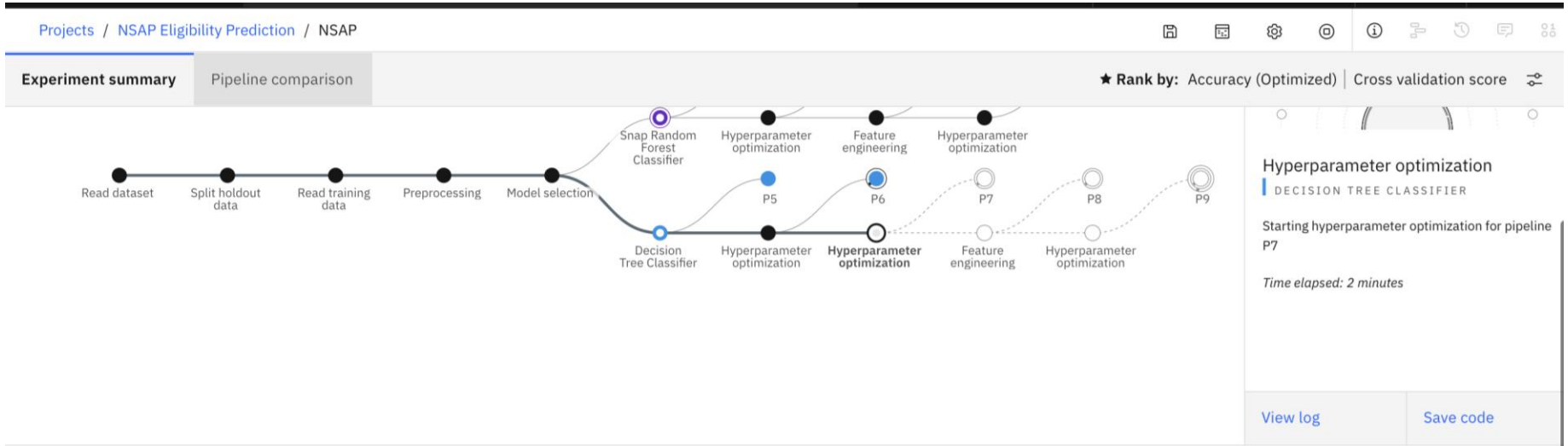
SYSTEM APPROACH

- The "System Approach" section outlines the overall strategy and methodology for developing and implementing the *Predicting Eligibility for NSAP using Machine Learning* project. Here's a suggested structure for this section:
- **System Requirements :**IBM Cloud (Lite Plan) ,Watsonx.ai Studio, chrome(browser), AI Kosh NSAP Dataset
- **Libraries and Tools Used:** IBM Watsonx.ai Auto AI, IBM Cloud Object Storage, Snap ML, Scikit-learn, Gradient boosting algorithm , Auto AI Leaderboard

ALGORITHM & DEPLOYMENT

- In the Algorithm section, describe the machine learning algorithm chosen for predicting NSAP scheme eligibility. Here's an example structure for this section:
- Algorithm Selection:
  - The Snap Random Forest Classifier was automatically selected by IBM Watsonx.ai Auto AI as the best-performing model due to its high accuracy, robustness in handling categorical features, and efficiency in processing structured demographic data relevant to scheme prediction.
- Data Input:
  - The model uses input features such as gender-wise beneficiary counts (total male, total female, total transgender), caste-wise distribution (Scheduled Caste, Scheduled Tribe, OBC, General), Aadhaar and mobile linkage counts, and regional identifiers like state and district names
- Training Process:
  - Auto AI handled the entire training workflow, including data preprocessing, automatic feature engineering, categorical encoding, hyperparameter tuning (HPO-1, HPO-2), and cross-validation. Models were trained and evaluated on a stratified split to ensure balanced performance.
- Prediction Process:
  - The final model classifies each data record (district) into its corresponding NSAP scheme (`scheme code`) based on the socio-economic profile. Predictions are made using the trained pipeline and can be extended for automated scheme eligibility verification if deployed

RESULT



Pipeline leaderboard

	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	00:00:34
	2	**Pipeline 3**	Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:27
	3	**Pipeline 2**	Snap Random Forest Classifier		0.981	HPO-1	00:00:07
	4	**Pipeline 1**	Snap Random Forest Classifier		0.981	*None*	00:00:02

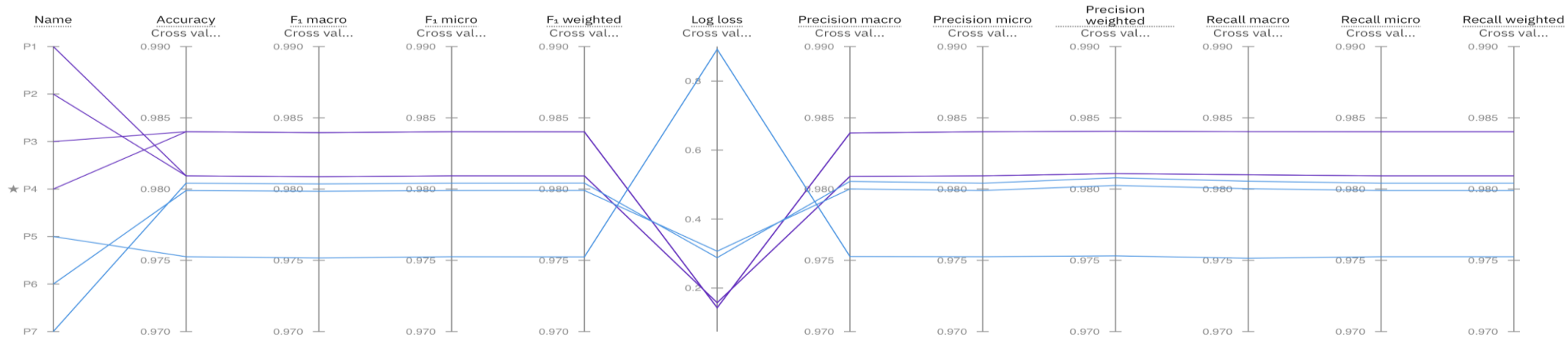
IBM **watsonx.ai Studio** | Search in your workspaces | Upgrade | Srigurunath Kudala's Account | London | SK

Projects / NSAP Eligibility Prediction / NSAP

Experiment summary | **Pipeline comparison** | ★ **Rank by:** Accuracy (Optimized) | Cross validation score

Metric chart

Prediction column: schemecode

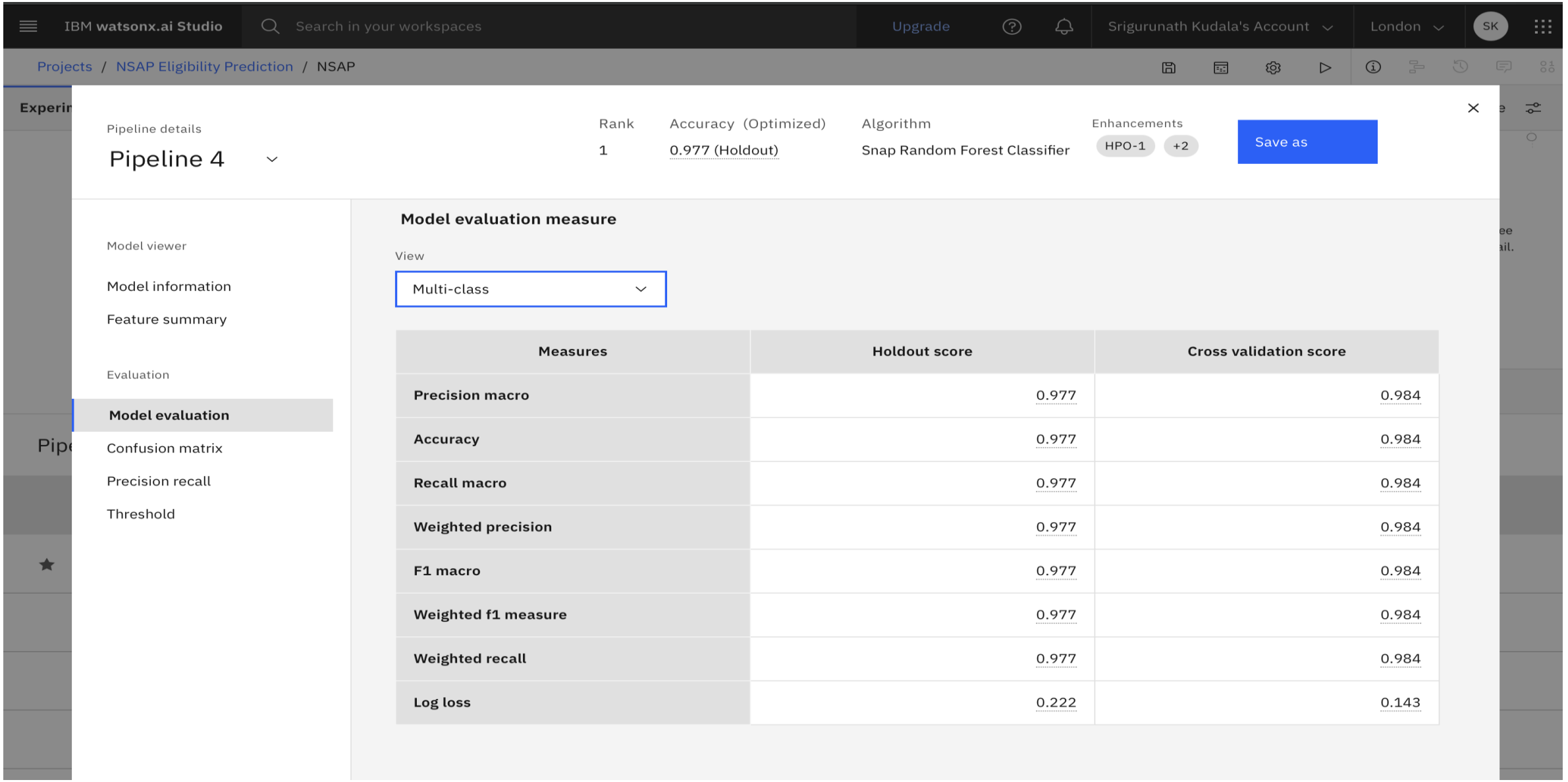


Pipeline leaderboard

	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	00:00:34
	2	**Pipeline 3**	Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:27

Pipeline comparison

edunet foundation



Measures	Holdout score	Cross validation score
Precision macro	0.977	0.984
Accuracy	0.977	0.984
Recall macro	0.977	0.984
Weighted precision	0.977	0.984
F1 macro	0.977	0.984
Weighted f1 measure	0.977	0.984
Weighted recall	0.977	0.984
Log loss	0.222	0.143

IBM watsonx.ai Studio | Search in your workspaces | Upgrade | Srigurunath Kudala's Account | London | SK

Projects / NSAP Eligibility Prediction / NSAP

Pipeline details

Pipeline 4

Rank	Accuracy (Optimized)	Algorithm	Enhancements
1	0.977 (Holdout)	Snap Random Forest Classifier	HPO-1 +2

Save as

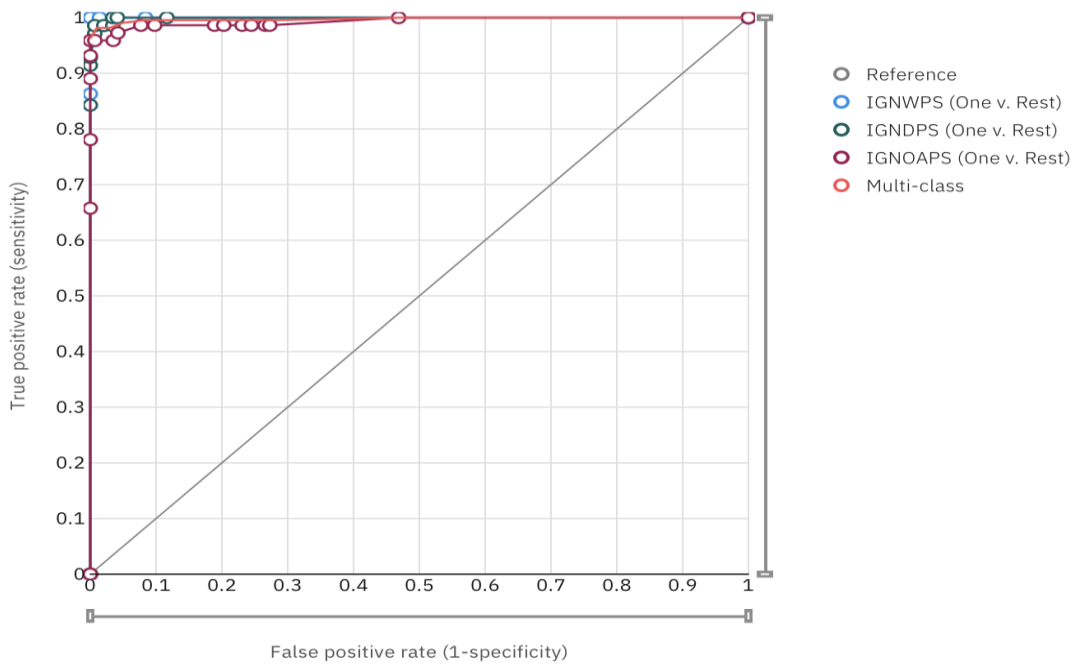
Model viewer

- Model information
- Feature summary

Evaluation

- **Model evaluation**
- Confusion matrix
- Precision recall
- Threshold

ROC curve



edunet foundation

CONCLUSION

- This project successfully applied IBM Watsonx.ai AutoAI to build a no-code machine learning model for predicting NSAP scheme eligibility. The Snap Random Forest Classifier delivered outstanding performance with **98.4% accuracy**, validating the effectiveness of automated ML pipelines. The solution streamlined classification using district-level socio-economic data, reducing manual effort and enhancing accuracy. By leveraging AI, the project demonstrates a scalable and reliable approach to improve transparency, efficiency, and fairness in public welfare distribution

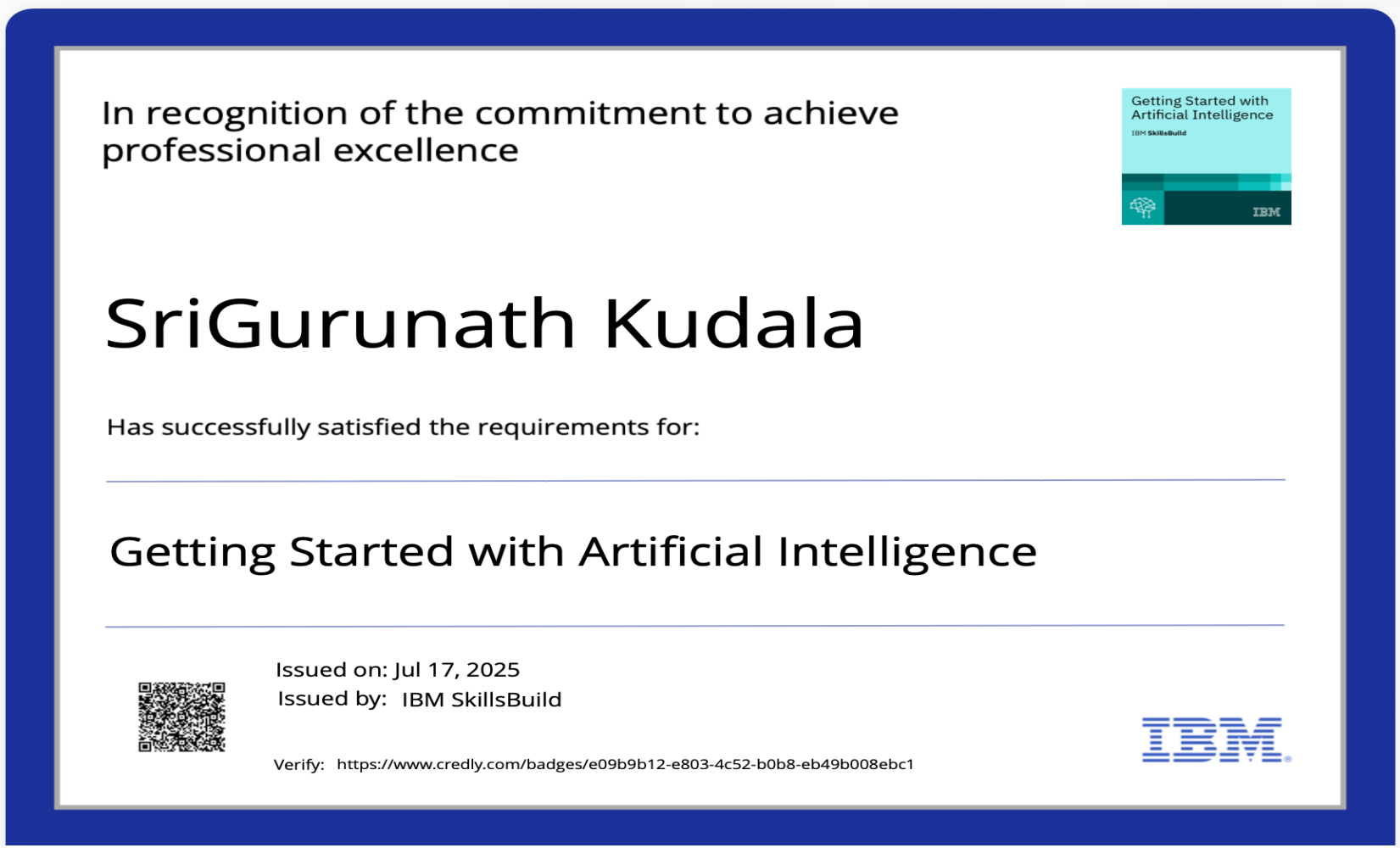
FUTURE SCOPE

- The system can be enhanced by integrating **individual-level applicant data** such as age, income, disability status, and family background for more personalized scheme predictions.
  Real-time data from government sources like **Aadhaar, SECC**, and **e-KYC** can be incorporated to improve accuracy and eliminate duplication.
  Model performance can be further optimized by experimenting with **advanced algorithms** or custom hyper parameter tuning beyond AutoAI.
  The solution can be scaled to support **state-wise or nationwide deployment**, covering diverse districts and beneficiary categories.
  Future iterations may also include **explainable AI**, **edge deployment**, or integration with **mobile/web portals** for on-the-spot eligibility checks by government officers.

REFERENCES

- **Using Algorithms to Match Eligible People to Social Benefits in India** – Policy Brief 🔗 https://www.idinsight.org/article/using-algorithms-to-match-elibible-people-to-social-benefits-in-india.
- **Leveraging Artificial Intelligence and Machine Learning to Reform India's Social Protection Framework** — Rajeshwar & Valaboju (2025) https://www.rjpn.org/ijcspub/papers/IJCSP25B1143.pdf
- **An Intelligent Social Protection Beneficiary Selection Scheme Using ML and a Mobile Application** — Alam & Chowdhury (2024) https://ideas.repec.org/a/ids/injdan/v16y2024i2p181-206.html
- **Predicting Poverty with Machine Learning and Geospatial Data** — Daoud et al. (2025) https://link.springer.com/chapter/10.1007/978-981-96-2544-4_4

IBM CERTIFICATIONS



IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

SriGurunath Kudala

for the completion of

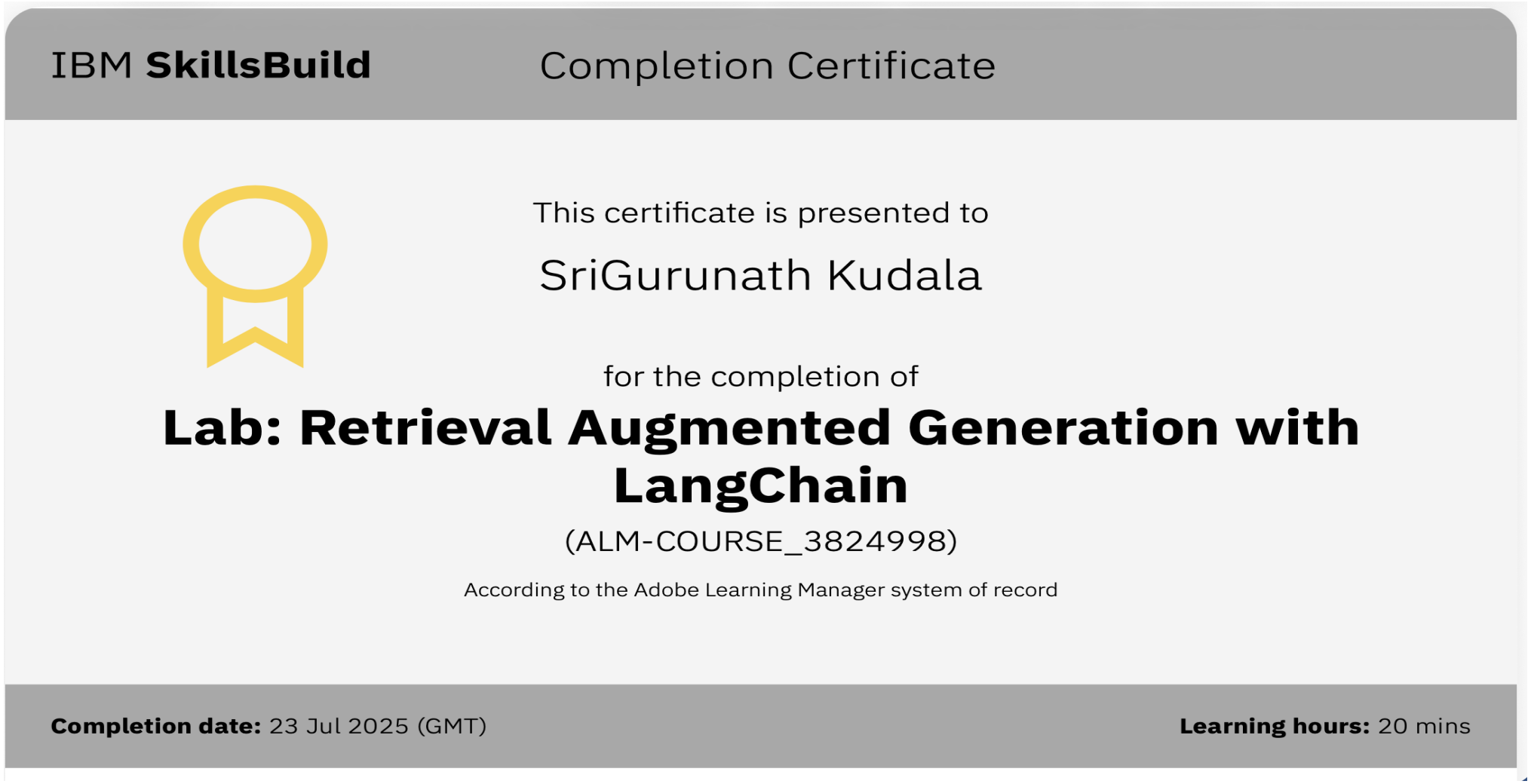
Journey to Cloud: Envisioning Your Solution

(PLAN-32CB1E21D8B4)

According to the Your Learning Builder - Plans system of record

Completion date: 17 Jul 2025 (GMT)

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

SriGurunath Kudala

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