

---

# **CAPSTONE PROJECT**

## **PREDICTING NSAP SCHEME ELIGIBILITY USING MACHINE LEARNING ON IBM CLOUD**

**Presented By:**

**1. Sivaranjany S S- Mepco Schlenk Engineering College - CSE**

# OUTLINE

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

# PROBLEM STATEMENT

The National Social Assistance Programme (NSAP) is a critical social welfare initiative by the Government of India that offers financial support to vulnerable sections such as elderly persons, widows, and individuals with disabilities from BPL households. Currently, the process of verifying eligibility and assigning the correct pension scheme is manual and time-consuming. This often results in delays, misallocation of benefits, and reduced accessibility for the intended beneficiaries.

# PROPOSED SOLUTION

- **Data Collection:**
  - District-wise NSAP beneficiary data was collected from the AI Kosh portal.
  - It includes demographic details, caste categories, and Aadhaar/mobile linkage counts.
- **Data Preprocessing:**
  - The dataset was cleaned for missing values and categorical fields were encoded.
  - Numerical features were normalized to enhance model performance.
- **Machine Learning Algorithm:**
  - IBM AutoAI selected and trained classification models, mainly using Snap Random Forest.
  - Feature engineering and hyperparameter tuning were applied automatically.
- **Deployment:**
  - The best-performing model was deployed using IBM Watson Machine Learning.
  - A REST API was generated for real-time predictions via UI or CSV uploads.
- **Evaluation:**
  - AutoAI evaluated pipelines using cross-validation accuracy.
  - The final model achieved 98.4% accuracy with optimized features.
- **Result**
  - The deployed model accurately predicts NSAP schemes based on input data.
  - It supports real-time or bulk testing and simplifies benefit allocation.

# SYSTEM APPROACH

- **System requirements**

- IBM Cloud Account

- WatsonStudio

- IBM AutoAI or Jupyter Notebook

- IBM Watson Machine Learning

- Python Libraries: pandas, scikit-learn, matplotlib

- **Library required**

- pandas for data loading and preprocessing

- sklearn for model building

- AutoAI for no-code model experimentation

- **Steps Followed:**

- Download dataset from AI Kosh (NSAP pension data)

- Upload it to IBM Watson Studio

- Launch AutoAI experiment

- Select schemecode as target column

- Evaluate best model and deploy

# ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

The Snap Random Forest Classifier was automatically selected and finalized by AutoAI based on **performance metrics, model stability, and cross-validation accuracy**. It was then deployed for real-time prediction through IBM Watson Machine Learning.

- **Data Input:**

- totalmale, totalfemale, totalsc, totalst, totalobc, totalaadhaar etc.

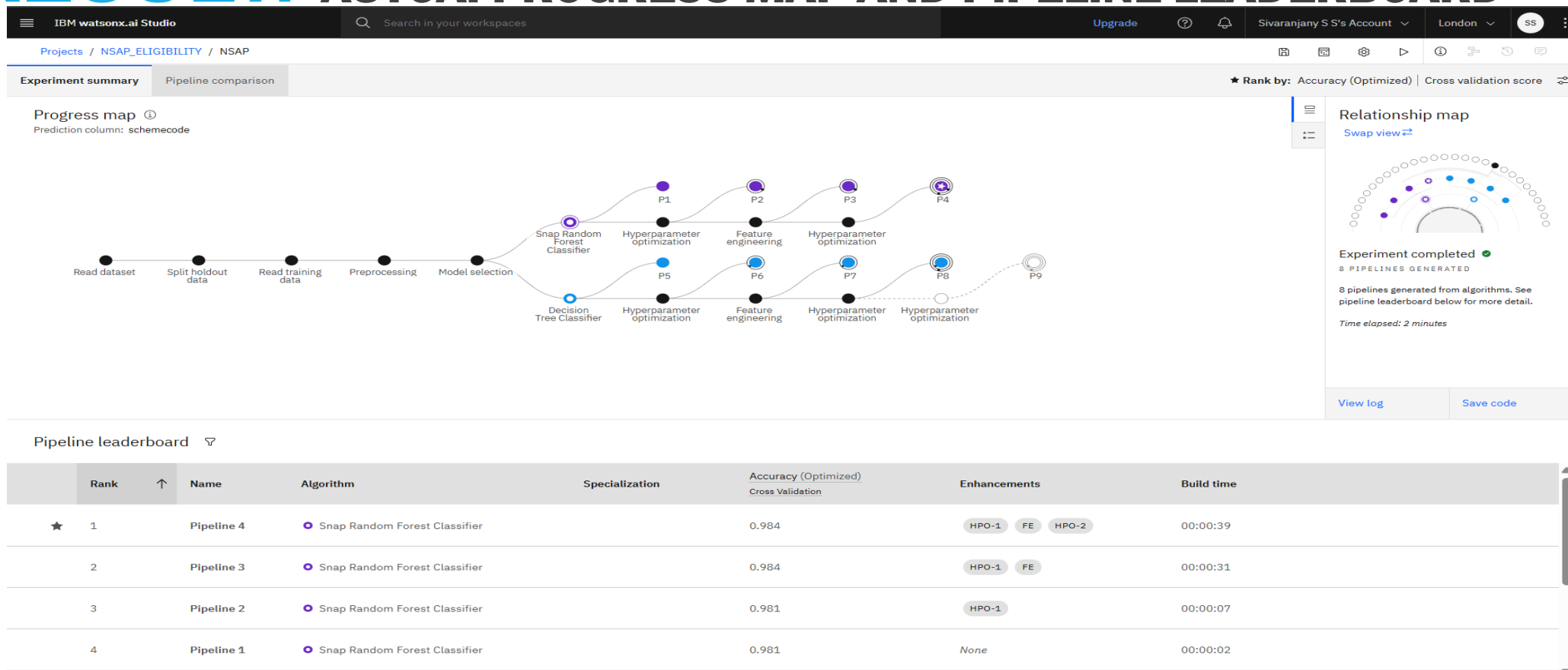
- **Training Process:**

- 80-20 train-test split
- Automatic hyperparameter tuning using AutoAI
- Model trained to classify one of the NSAP schemes

- **Prediction Process:**

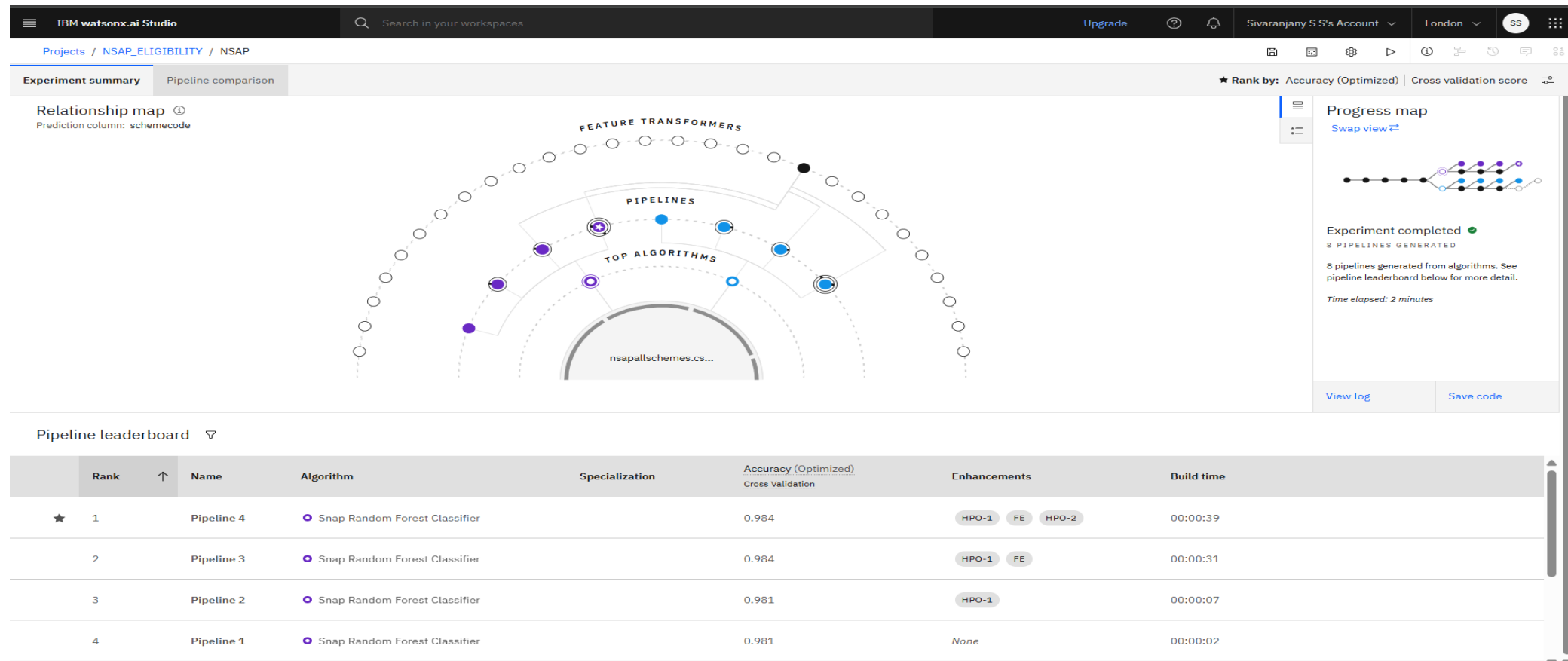
- Model deployed using Watson Machine Learning.
- API endpoint generated for live prediction.
- Can accept new applicant data and return predicted scheme

# RESULT: AUTOAI PROGRESS MAP AND PIPELINE LEADERBOARD



AutoAI generated 8 pipelines using various ML algorithms and enhancements. The best-performing pipeline (Pipeline 4) used the Snap Random Forest Classifier and achieved an optimized accuracy of 98.4%. Enhancements like feature engineering (FE) and hyperparameter optimization (HPO) were automatically applied to improve performance.

# RESULT : RELATIONSHIP MAP AND PIPELINE ACCURACY



The Relationship Map shows how AutoAI explored various combinations of algorithms, feature transformers, and tuning strategies.

The system automatically selected and ranked the pipelines based on cross-validation accuracy.

The top 4 pipelines used Random Forest as the base algorithm with different levels of enhancement.



# RESULT : MODEL DEPLOYMENT AND TESTING

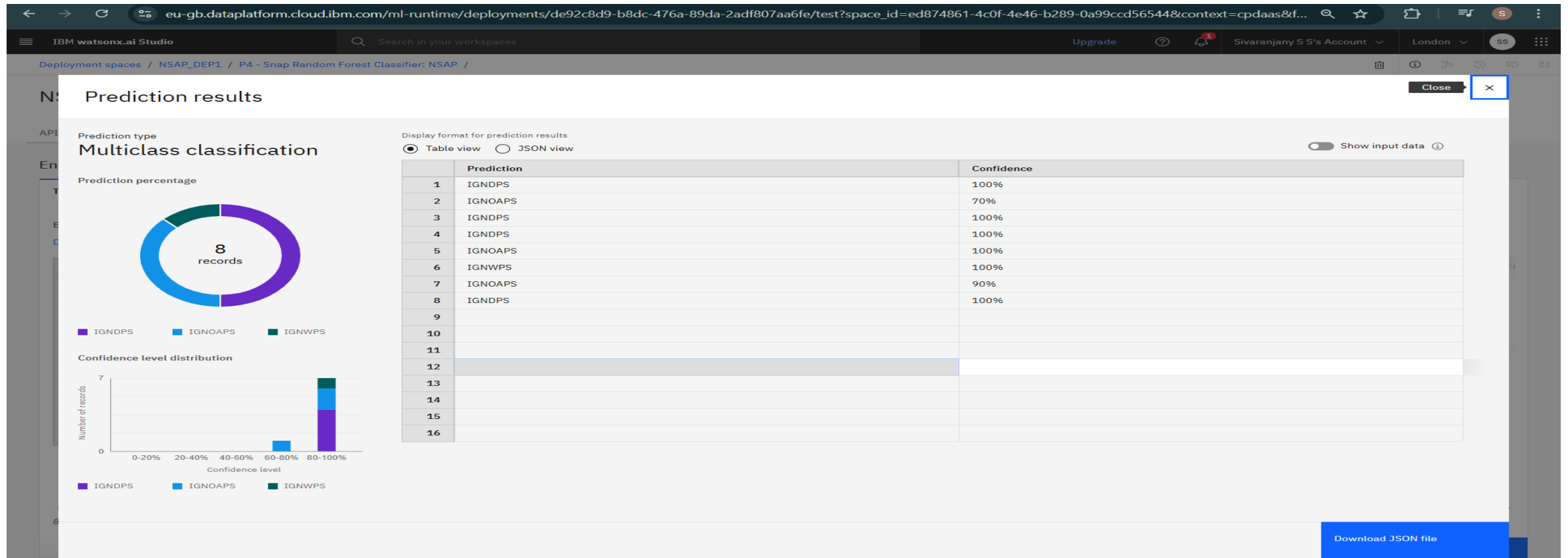
The screenshot shows the IBM Watson Machine Learning web interface for a deployed model named 'NSAP\_eligibility'. The interface includes a navigation bar with the IBM logo, a search bar, and user account information. The main content area is titled 'NSAP\_eligibility' and shows the model is 'Deployed' and 'Online'. Below this, there are tabs for 'API reference' and 'Test'. The 'Test' tab is active, showing an 'Enter input data' section. This section has two tabs: 'Text' and 'JSON'. The 'Text' tab is selected, and it displays a table with 8 rows of input data. The table has 11 columns: 'fnyear (other)', 'lgdstatecode (double)', 'statename (other)', 'lgddistrictcode (double)', 'districtname (other)', 'totalbeneficiaries (double)', 'totalmale (double)', 'totalfemale (double)', 'totaltransgender (double)', 'totalsc (double)', and 'total'. The data is as follows:

	fnyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)	totaltransgender (double)	totalsc (double)	total
1	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	108	72	36	0	0	3
2	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	8438	5059	3379	0	37	235
3	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	202	100	102	0	1	15
4	2025-2026	1	JAMMU AND KASH	10	POONCH	310	211	99	0	0	77
5	2025-2026	1	MANIPUR	10	THOUBAL	5958	3958	2000	0	2	1347
6	2025-2026	1	ASSAM	10	BAKSA	280	0	280	10	2	21
7	2025-2026	1	JAMMU AND KASH	5	BARAMULLA	7123	4201	2922	0	10	304
8	2025-2026	1	BIHAR	5	BHOJPUR	196	128	68	0	0	5
9											
10											

Below the table, there is a scroll bar and the text '8 rows, 15 columns'. At the bottom right of the 'Test' tab, there is a blue button labeled 'Predict'.

The deployed model is tested with 8 real-world input records using IBM Watson Machine Learning's web UI. The prediction can be triggered with a single click, and results are returned instantly.

# RESULT : DEPLOYED MODEL – TEST INPUT INTERFACE



After training, the model was successfully deployed using Watson ML with an online endpoint. The interface allows uploading or pasting new applicant data to get real-time predictions. The system supports CSV uploads (max 50MB) and displays the prediction results upon clicking “Predict.”

# CONCLUSION

- The machine learning model built using IBM Watson Studio successfully predicts the appropriate NSAP scheme based on applicant data.  
It demonstrates strong performance, simplifies the verification process, and supports the automation of social scheme allocation.  
This AI-based approach can help the government make the delivery of social benefits more efficient and equitable.

# FUTURE SCOPE

- Integrate with government databases for real-time eligibility checking. Extend the model with additional data like disability percentage, income levels, etc. Build a chatbot interface for beneficiaries to check their scheme eligibility. Enhance model explainability using SHAP or LIME for transparency.

---

# REFERENCES

- **AI Kosh Dataset:** <https://aikosh.indiaai.gov.in>
- **IBM Watson Studio:** <https://cloud.ibm.com>
- **Scikit-learn Documentation:** <https://scikit-learn.org>

# IBM CERTIFICATIONS

In recognition of the commitment to achieve  
professional excellence



Sivaranjany S S

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/28e3de49-ffdc-4bb9-954e-73d3007b78b9>



# IBM CERTIFICATIONS

In recognition of the commitment to achieve  
professional excellence



## Sivaranjany S S

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/3955e757-e9a6-48fa-a6e9-4026f84730bb>



# IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to

Sivaranjany S S

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





**THANK YOU**