CAPSTONE PROJECT

PREDICTING ELIGIBILITY FOR NSAP USING MACHINE LEARNING

Presented By:

1. Name: FAISHAL RAHMAN ANSARI
2. College Name: Sambhram Institute of Technology
3. Dept Name: Computer Science Engineering

edunet foundation

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

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 can be a time consuming and error-prone process. Delays or incorrect allocation can prevent deserving individuals from receiving timely financial aid. Your task is to design, build, and evaluate a multi-class classification model that can accurately predict 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 aims to streamline the allocation of NSAP schemes by accurately predicting the most suitable scheme for each applicant. By utilizing data analytics and machine learning techniques, the solution will enhance the efficiency and fairness of scheme distribution. The key components of the solution are as follows:
- Data Collection:
  - Collect historical NSAP applicant data, including demographic attributes such as gender, caste category, state, and district from official beneficiary records.
  - Gather socio-economic information like Aadhaar linkage status, mobile number availability, and total number of beneficiaries per category (e.g., SC/ST/Gen/OBC) to support eligibility prediction.
- Data Preprocessing:
  - Handle missing values, duplicates, and inconsistencies in applicant-related data such as gender, caste category, and mobile/Aadhaar availability.
  - Perform feature engineering to derive impactful indicators like beneficiary count trends, Aadhaar coverage ratio, or scheme-wise gender distribution, which influence eligibility.
- Machine Learning Algorithm:
  - Use algorithms like Random Forest, XGBoost, or Neural Networks to classify applicants into the correct NSAP scheme based on their profile.
  - Leverage key demographic and socio-economic features (e.g., gender, caste, Aadhaar status) to improve the model's prediction accuracy and fairness.
- Deployment:
  - Develop a user-friendly interface or web-based application where government officials can input applicant data and instantly receive predicted NSAP scheme suggestions.
  - Deploy the solution on a scalable and secure platform like IBM Cloud or Watson Studio, ensuring smooth performance, data privacy, and accessibility for use at scale across departments.
- Evaluation:
  - Evaluate model performance using metrics like accuracy, precision, recall, and F1-score.
  - Use confusion matrix and class-wise analysis to detect misclassifications and fine-tune the model with new data.

SYSTEM APPROACH

The "System Approach" outlines the overall strategy and methodology for building and deploying the NSAP scheme prediction model to assist government agencies in automating applicant classification.

SYSTEM REQUIREMENTS

- **A dataset containing demographic and socio-economic data of applicants along with their previously assigned NSAP scheme codes.**
- **IBM Cloud or Watson Studio environment for model development and deployment.**
- **Web or mobile interface to input applicant data and receive real-time predictions.**

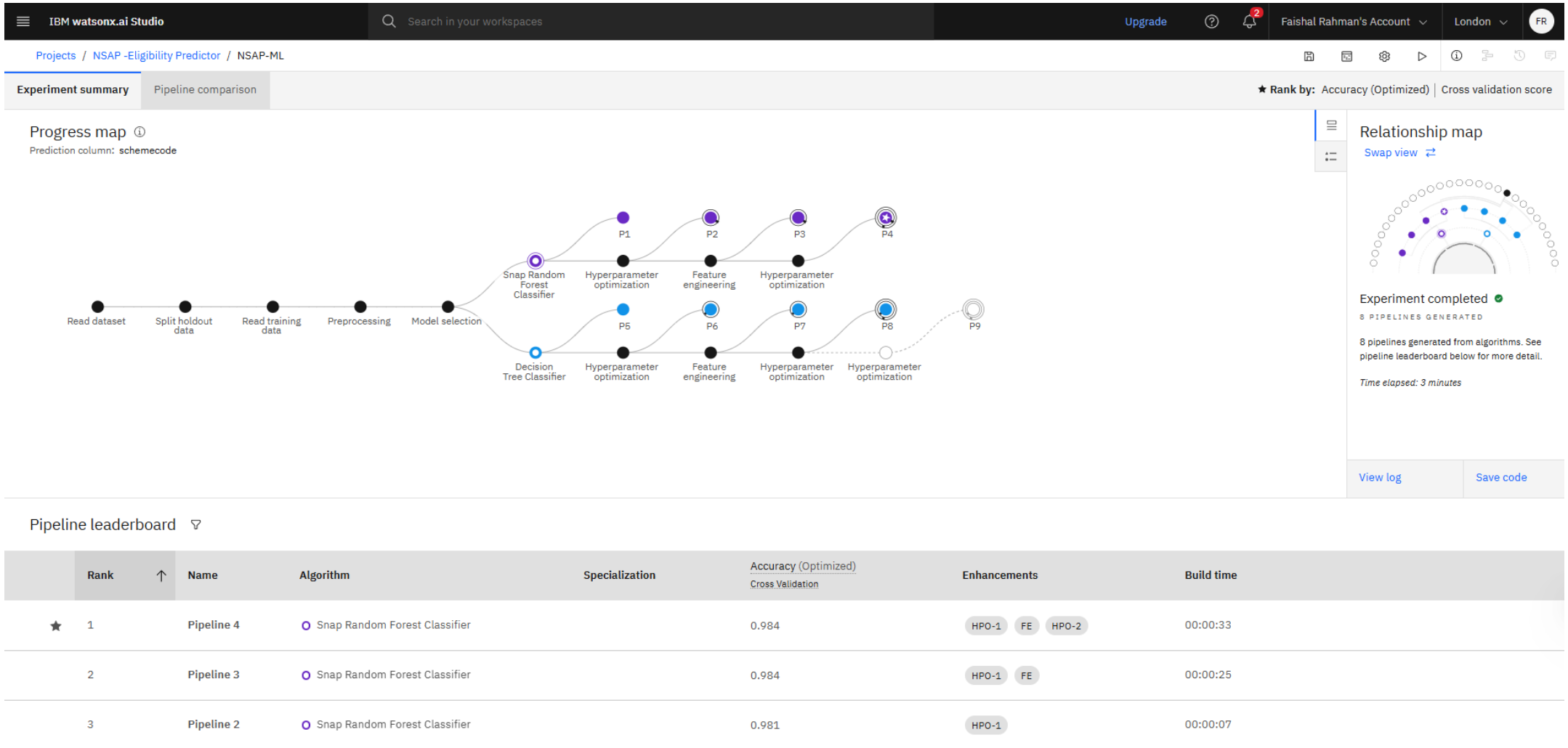
LIBRARY REQUIRED TO BUILD THE MODEL

- **Automated Data Preprocessing:** AutoAI handles missing values, categorical encoding, and feature engineering internally.
- **Model Selection & Optimization:** It automatically selects the best classification algorithms (e.g., Decision Trees, Random Forest, XGBoost) based on data patterns.
- **Pipeline Generation:** Multiple model pipelines are generated, ranked by evaluation metrics such as Accuracy and F1-Score.
- **One-Click Deployment:** The best-performing pipeline is deployed as a web service on IBM Cloud, ready for integration into applications.

ALGORITHM & DEPLOYMENT

- In the Algorithm section, describe the machine learning algorithm chosen for predicting bike counts. Here's an example structure for this section:
- **Algorithm Selection:**
  - IBM Watson AutoAI evaluates multiple classification algorithms including Decision Trees, Random Forest, and XGBoost, automatically selecting the one that best fits the patterns within the applicant dataset.
  - The selection is driven by data distribution, feature types, and the complexity of multi-class classification, ensuring high generalization and robustness.
- **Data Input:**
  - Input features include demographic and socio-economic details such as gender, caste category, Aadhaar possession, and mobile access. These attributes help determine the eligibility and most suitable scheme for each applicant.
- **Training Process:**
  - AutoAI automatically preprocesses the dataset, performs feature engineering, and applies cross-validation to train multiple models. It then ranks them based on evaluation metrics like accuracy and F1-score.
- **Prediction Process:**
  - The best-performing model is deployed and used to predict scheme codes for new applicants in real-time. This enables fast, consistent, and transparent decision-making for government officials.

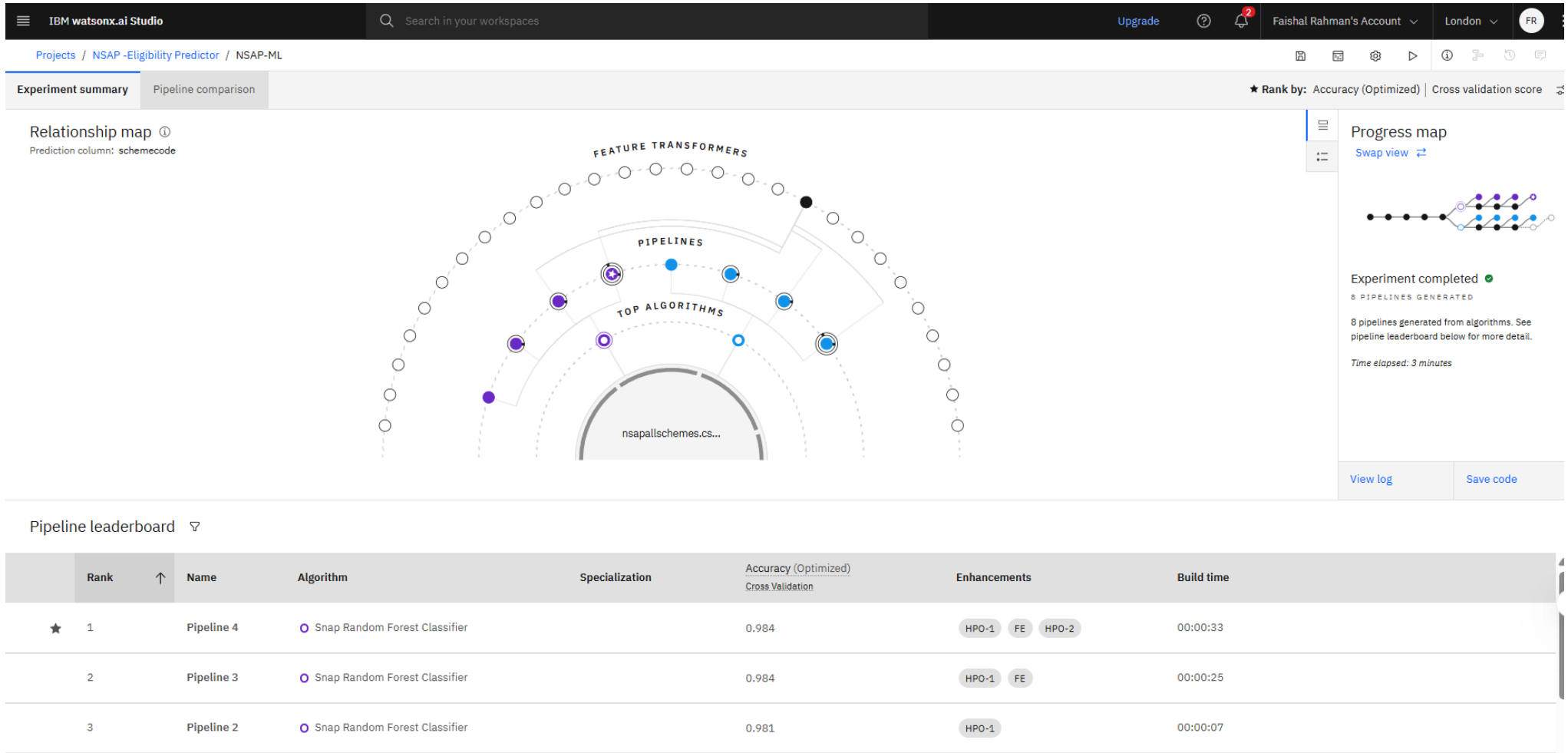
RESULT



Pipeline leaderboard

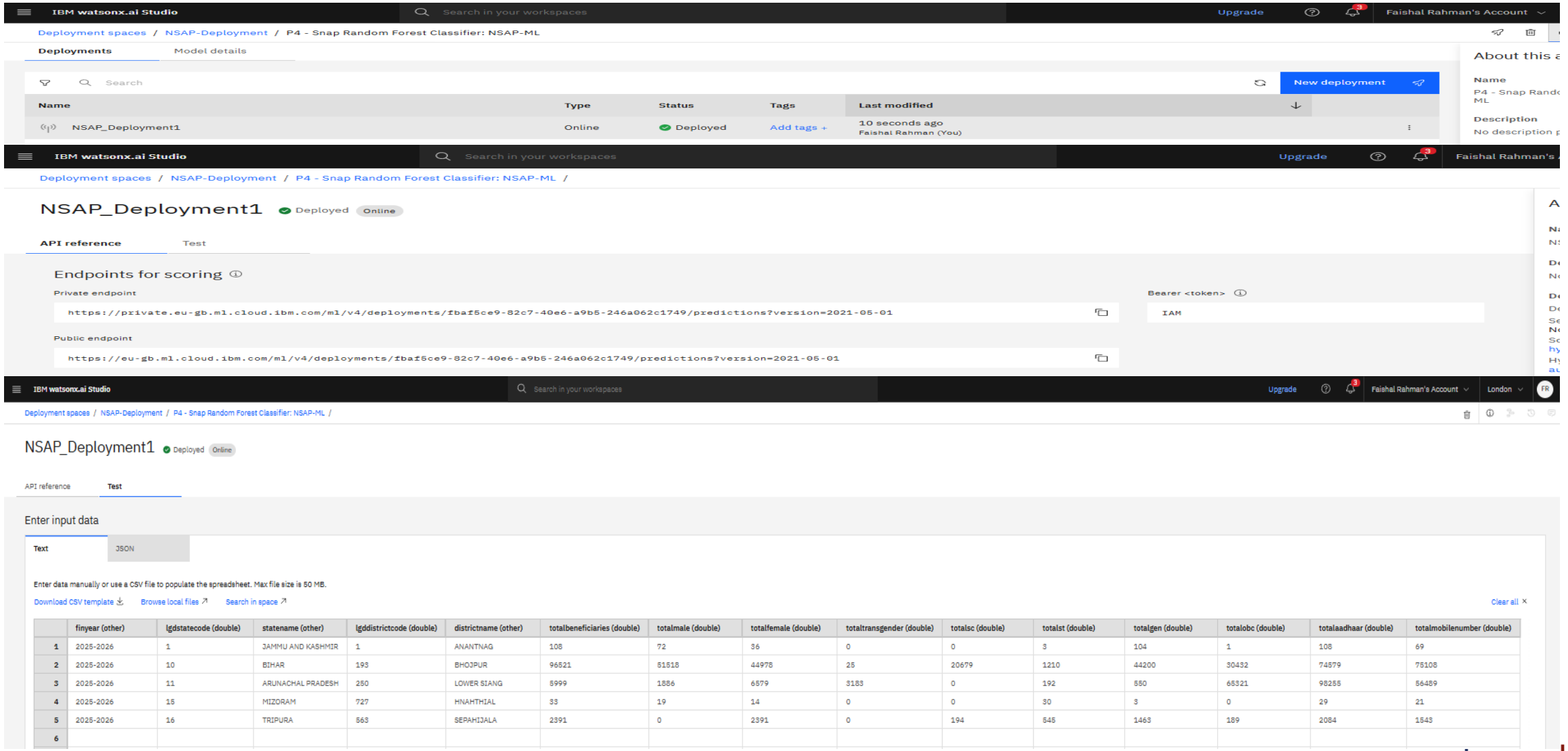
	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 4	Snap Random Forest Classifier		0.984	HPO-1 FE HPO-2	00:00:33
	2	Pipeline 3	Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:25
	3	Pipeline 2	Snap Random Forest Classifier		0.981	HPO-1	00:00:07

RESULT



edunet foundation

RESULT



	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)	totaltransgender (double)	totalsc (double)	totalst (double)	totalgen (double)	totalobc (double)	totalaadhaar (double)	totalmobilenumber (double)
1	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	108	72	36	0	0	3	104	1	108	69
2	2025-2026	10	BIHAR	193	BHOJPUR	96521	51518	44978	25	20679	1210	44200	30432	74579	75108
3	2025-2026	11	ARUNACHAL PRADESH	250	LOWER SIANG	5999	1886	6579	3183	0	192	550	65321	98255	56489
4	2025-2026	15	MIZORAM	727	HNAHTHIAL	33	19	14	0	0	30	3	0	29	21
5	2025-2026	16	TRIPURA	563	SEPAHIJALA	2391	0	2391	0	194	545	1463	189	2084	1543
6															

edunet foundation

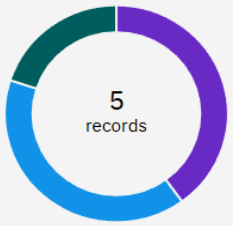
RESULT

Prediction results

Prediction type

Multiclass classification

Prediction percentage



IGNDPS IGNOAPS IGNWPS

Confidence level distribution

IGNDPS IGNOAPS IGNWPS

Display format for prediction results

(•) Table view ( ) JSON view

Show input data

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

Download JSON file

edunet foundation

CONCLUSION

- The proposed solution effectively leverages IBM Watson Studio's AutoAI to automate the classification of NSAP schemes based on applicant demographic and socio-economic data. The no-code, scalable approach ensures faster, accurate, and fair scheme assignment, supporting timely delivery of benefits.
- During implementation, challenges such as data imbalance and feature relevance were addressed through AutoAI's automated preprocessing and model selection. Future improvements may include incorporating additional contextual data (e.g., local economic indicators) to further enhance prediction quality. This system supports the government's goal of transparency, efficiency, and inclusiveness in welfare scheme distribution.

FUTURE SCOPE

- The solution can be enhanced by integrating additional applicant-level data such as health indicators or family income to improve prediction accuracy. As the system scales, it can be expanded to support classification across other government welfare schemes beyond NSAP.
- Further improvements could include retraining the model with real-time data streams, incorporating explainable AI for transparency in decision-making, and integrating the solution with mobile apps or e-governance portals to enable on-the-spot eligibility assessments in rural and remote regions.

REFERENCES

•Government of India – Ministry of Rural Development, NSAP Guidelines and Scheme Detail

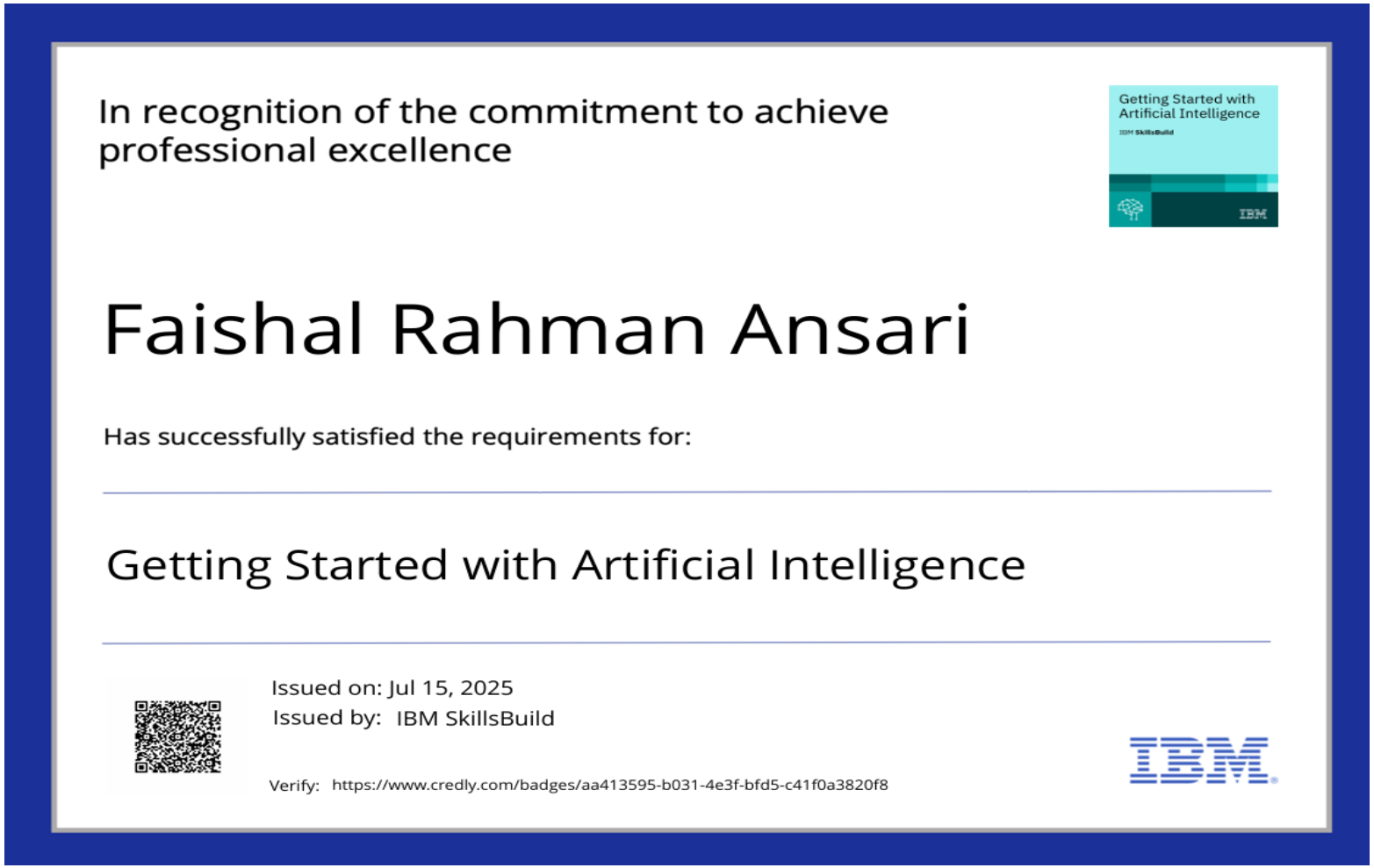
•IBM Watson Studio & AutoAI Documentation – https://www.ibm.com/cloud/watson-studio

•Scikit-learn Documentation for Classification Metrics – https://scikit-learn.org/stable/modules/model_evaluation.html

•XGBoost: A Scalable Tree Boosting System – Chen, T., & Guestrin, C. (2016)

•Research Articles on Automated Machine Learning (AutoML) for Social Welfare Programs

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Faishal Rahman Ansari

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/aa413595-b031-4e3f-bfd5-c41f0a3820f8

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

Faishal Rahman Ansari

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/3758d3ac-1e4f-43f9-aebe-bb317ecf844c

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Faishal Rahman Ansari

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

edunet foundation