CAPSTONE PROJECT

Predicting Eligibility for NSAP Schemes using Machine Learning

Presented By:
Bharanieeswaran R – KPR Institute of Engineering and Technology – CSE Department

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 Programme (NSAP) provides financial aid to vulnerable BPL individuals through sub-schemes like IGNOAPS, IGNWPS, and IGNDPS. Currently, scheme allocation is done manually, leading to delays and misclassification. This project aims to build a machine learning model that can automatically and accurately predict the appropriate NSAP scheme for an applicant based on demographic and socio-economic data, improving efficiency and ensuring timely benefit delivery.

Proposed Solution

- he proposed system aims to address the challenge of accurately classifying applicants under the National Social Assistance Programme (NSAP) into the appropriate welfare scheme. This involves leveraging machine learning techniques and demographic data to automate the decision-making process and ensure timely allocation of benefits. The solution will consist of the following components:
- Data Collection:
- Gather district-wise beneficiary data from the AI_KOSH dataset under NSAP. The dataset includes demographic and socio-economic information such as gender distribution, caste categories, Aadhaar linkage, and mobile number availability.
- Data Preprocessing:
- Clean and preprocess the collected data to handle missing values, inconsistencies, and outliers. Apply feature engineering techniques to extract meaningful attributes that influence scheme eligibility, such as SC/ST ratios or Aadhaar coverage.
- Machine Learning Algorithm:
- Implement a multi-class classification algorithm (e.g., Random Forest Classifier) to predict the suitable scheme (IGNOAPS, IGNWPS, or IGNDPS) for each applicant. The model will be trained on historical scheme allocation patterns using the processed data.
- Deployment:
- Develop a system that allows users to input applicant data manually or through CSV files. Deploy the trained model using IBM Cloud Lite services to enable scalable and accessible predictions for government officials.
- Evaluation:
- Evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score. Analyze confusion matrices and feature importance to assess model effectiveness. Continuously improve the model based on results and new data inputs.
- Result:
- The system successfully predicts the appropriate NSAP scheme for applicants, reducing manual effort and improving the speed and accuracy of benefit delivery.
-

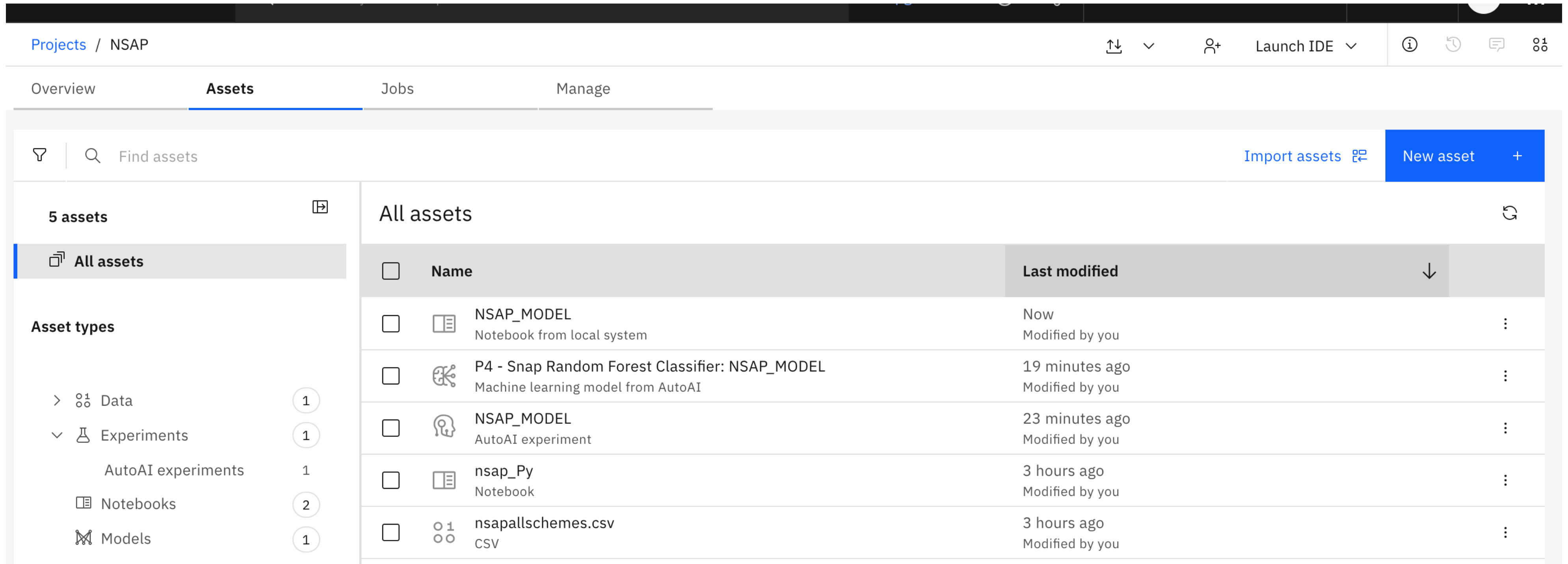
System Approach

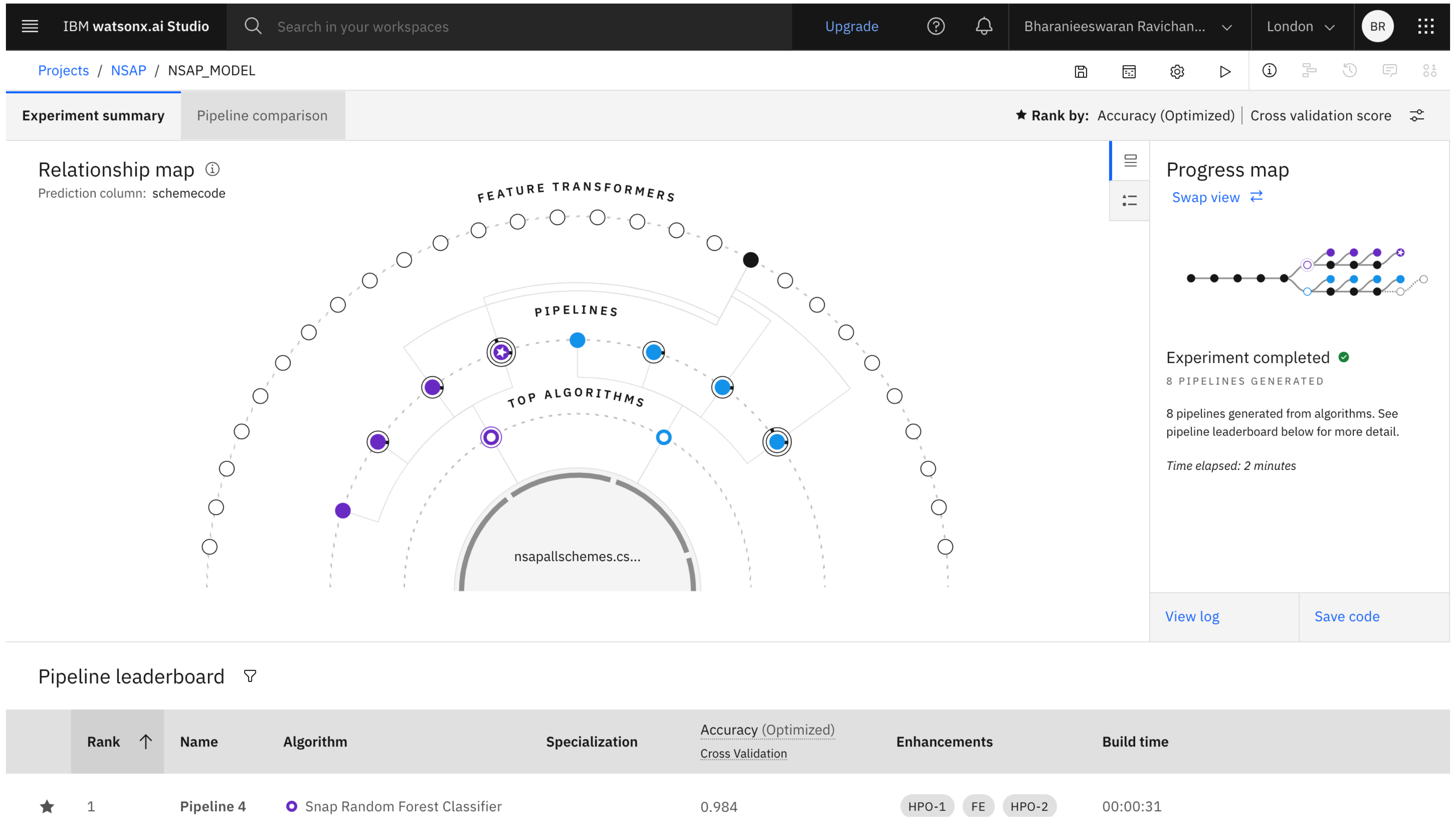
The "System Approach" section outlines the overall strategy and methodology used to develop and implement the NSAP scheme prediction system. This approach includes defining system requirements, selecting appropriate tools, and designing a workflow that ensures accurate and efficient prediction of the eligible NSAP scheme for each applicant

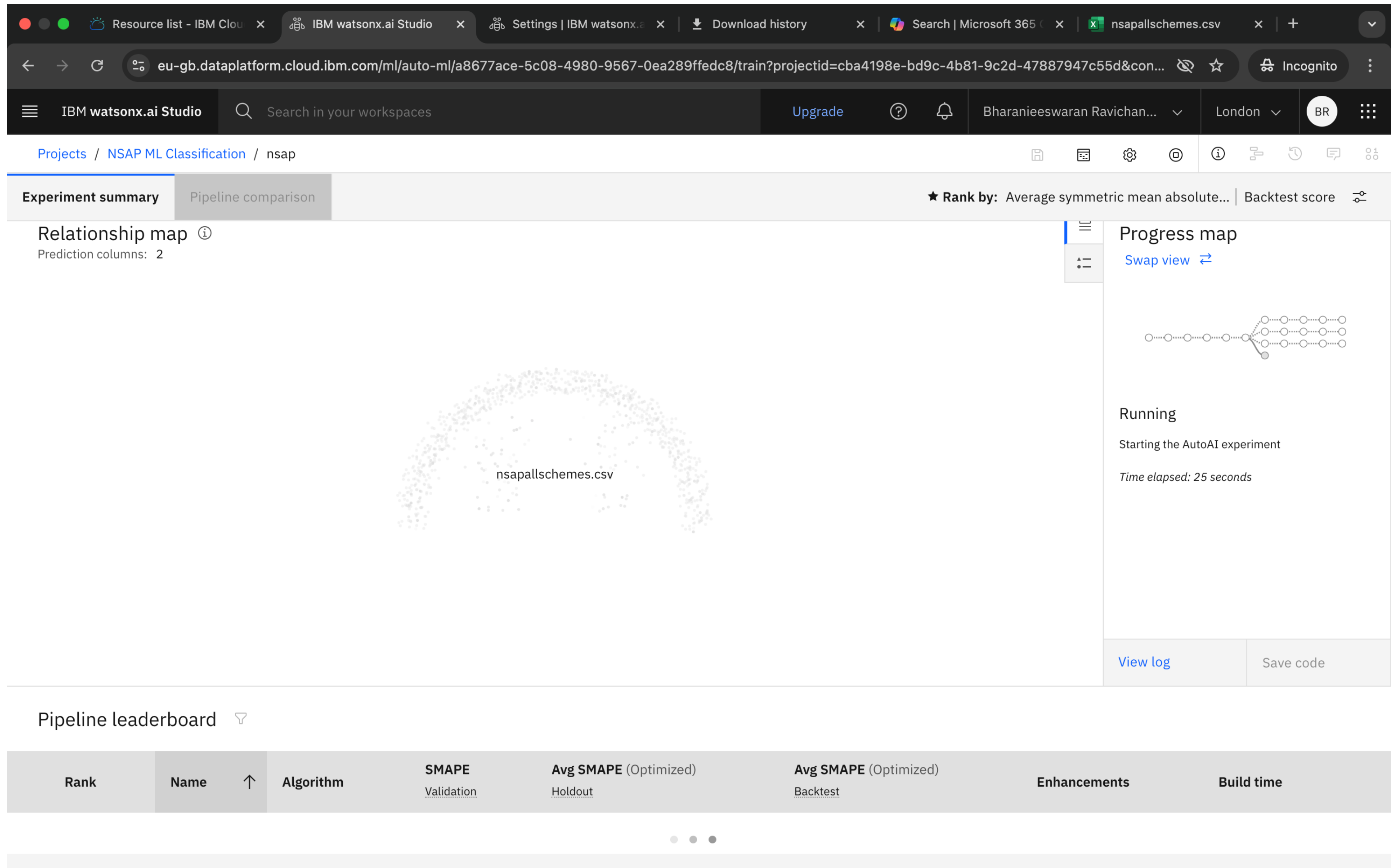
Algorithm & Deployment

- Algorithm Selection:
- For this project, the Random Forest Classifier was chosen as the primary machine learning algorithm. Random Forest is a robust ensemble-based classification model that works well with structured tabular data and handles both categorical and continuous features efficiently. It is suitable for multi-class classification problems, such as predicting which NSAP scheme (IGNOAPS, IGNWPS, IGNDPS) a beneficiary qualifies for, based on patterns in demographic and socio-economic data.
- Data Input:
- The model takes input features from the AI_KOSH dataset, including:
- Total male, female, and transgender beneficiaries
- SC, ST, OBC, and General category counts
- Aadhaar-linked and mobile-linked beneficiary counts
- Total beneficiaries in the district
- These features provide a socio-economic profile of each region, which influences eligibility under different NSAP schemes.
- Training Process:
- The dataset was preprocessed to remove irrelevant fields (like district names and financial year), and label encoding was applied to the target column (schemecode). The data was then split into training and testing sets. The Random Forest model was trained using the training data and evaluated on the test set. Hyperparameter tuning was done using grid search and cross-validation to optimize the number of estimators and tree depth.
- Prediction Process:
- Once trained, the model predicts the most suitable scheme code for a given set of input features. New data can be provided in CSV format or through an interface, and the model returns the predicted NSAP scheme. The prediction logic is deployed in a cloud environment (IBM Cloud Lite), allowing for real-time use by administrators or application portals.

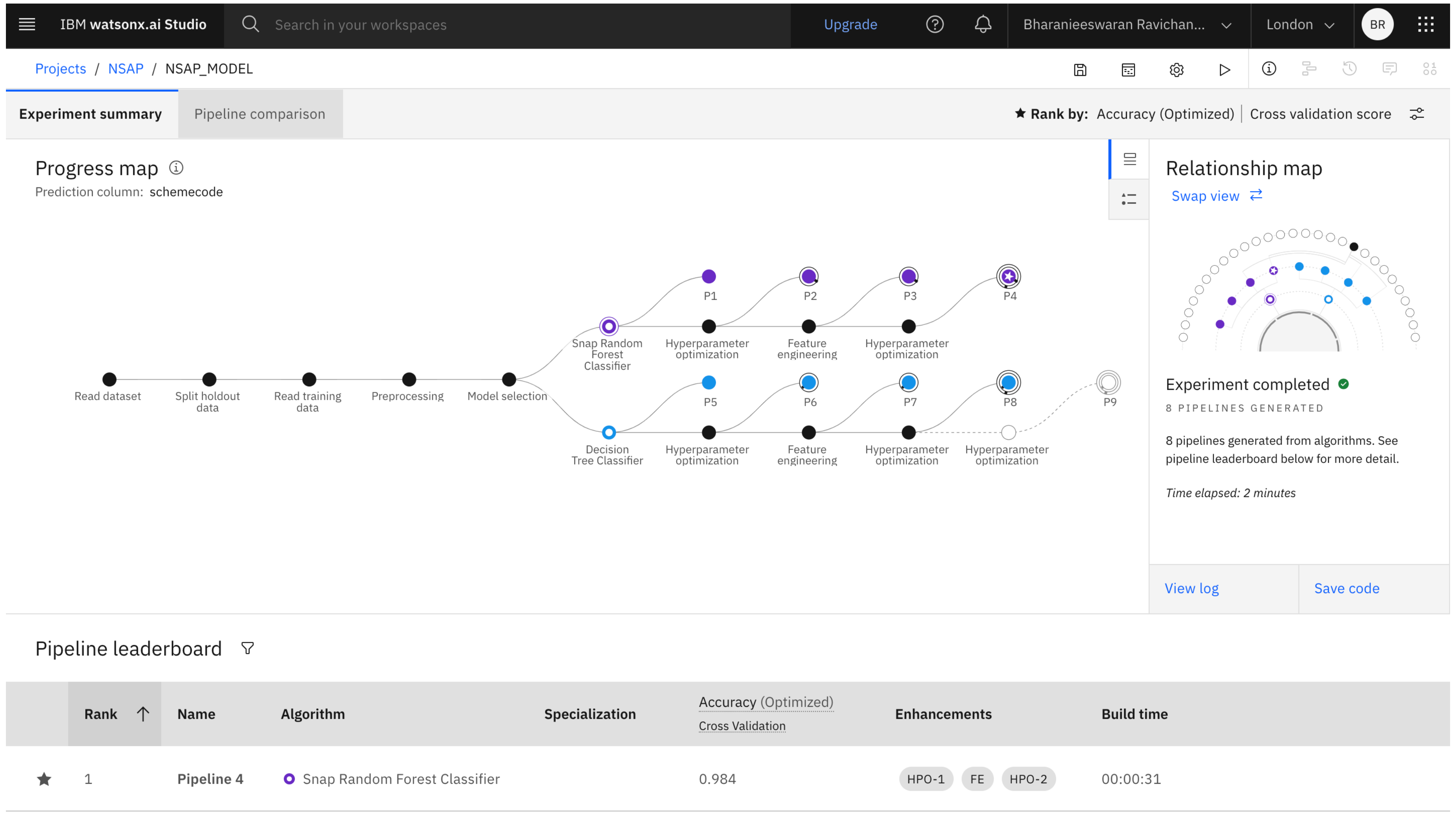
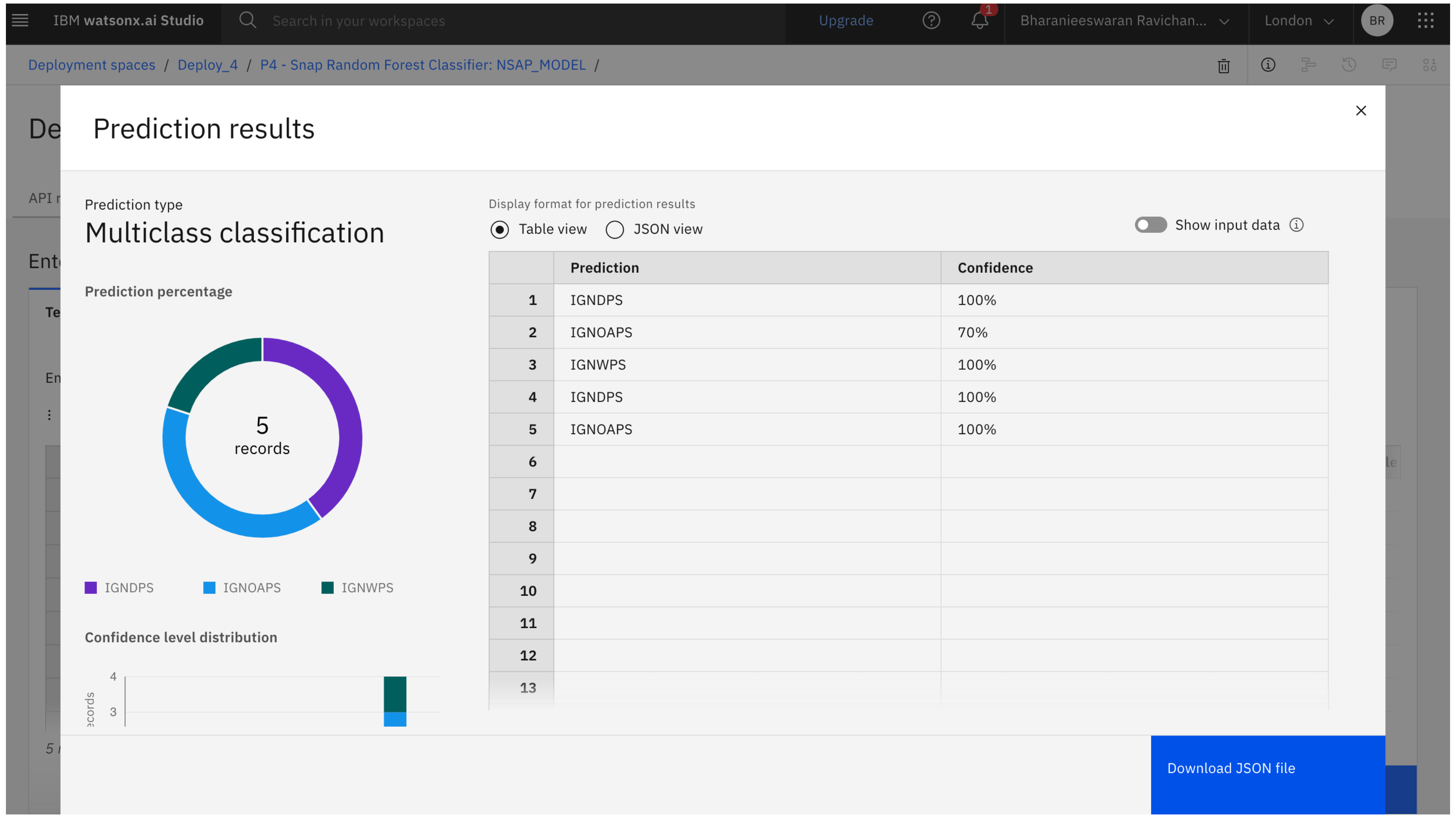
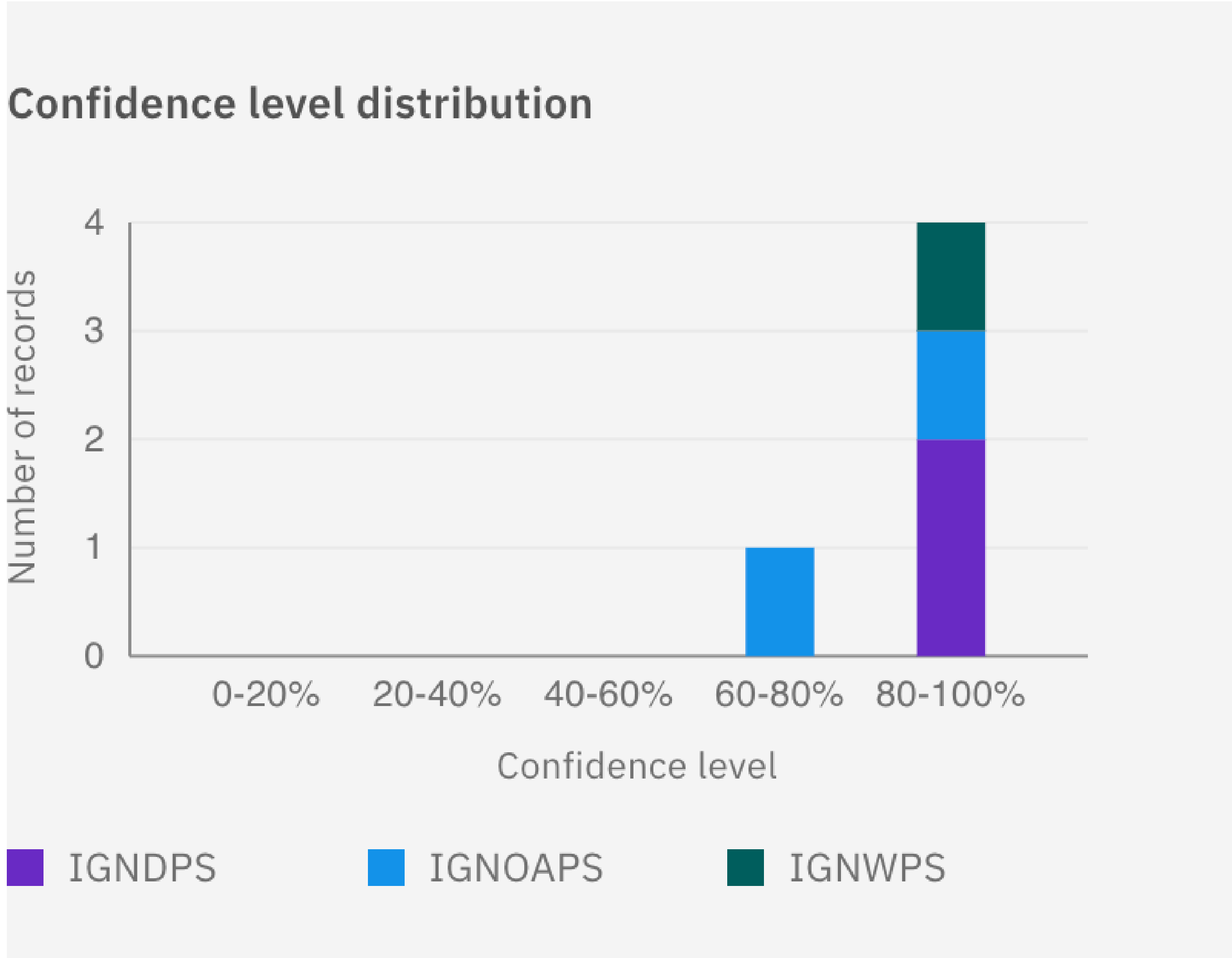
Result







edunet foundation


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Bharanieeswaran Ravichan...
London
Projects / NSAP / NSAP_MODEL
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Progress map
Prediction column: schemecode
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
Snap Random Forest Classifier
Decision Tree Classifier
Hyperparameter optimization
Feature engineering
P1
P2
P3
P4
P5
P6
P7
P8
P9
Relationship map
Swap view
Experiment completed
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 2 minutes
View log
Save code
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
00:00:31

IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Bharanieeswaran Ravichan...
London
Deployment spaces / Deploy_4 / P4 - Snap Random Forest Classifier: NSAP_MODEL /
Prediction results
Prediction type
Multiclass classification
Display format for prediction results
Table view
JSON view
Show input data
Prediction percentage
5
records
IGNDPS
IGNOAPS
IGNWPS
Confidence level distribution
Prediction
Confidence
1
IGNDPS
100%
2
IGNOAPS
70%
3
IGNWPS
100%
4
IGNDPS
100%
5
IGNOAPS
100%
Download JSON file

Confidence level distribution
Number of records
0
1
2
3
4
0-20%
20-40%
40-60%
60-80%
80-100%
Confidence level
IGNDPS
IGNOAPS
IGNWPS

Conclusion

- The findings of this project demonstrate that machine learning can effectively automate the classification of beneficiaries under the National Social Assistance Programme (NSAP). By using district-level demographic and socio-economic data, the Random Forest classification model was able to predict the appropriate NSAP scheme—IGNOAPS, IGNWPS, or IGNDPS—with high accuracy. This significantly reduces manual effort, minimizes human error, and ensures timely delivery of financial assistance to eligible individuals.
- During implementation, challenges included handling categorical variables, class imbalance, and ensuring that the model remained interpretable for decision-makers. These were addressed through data preprocessing, feature selection, and model evaluation techniques such as confusion matrices and feature importance analysis.
- The proposed solution proved to be efficient, scalable, and easy to integrate into government workflows using IBM Cloud Lite services. It can assist authorities in making quick and accurate eligibility decisions, ultimately improving the transparency and effectiveness of social welfare distribution.

Future scope

The current system demonstrates a promising approach for automating NSAP scheme classification using machine learning. However, there are several opportunities to enhance and expand its capabilities.

One key improvement involves incorporating individual-level applicant data rather than district-level aggregates. This would allow the system to make personalized predictions and significantly increase the model's practical applicability. Integrating real-time data feeds from digital application portals or Aadhaar-linked databases can also enhance prediction accuracy and reduce dependency on static datasets.

Another enhancement is the optimization of the algorithm using more advanced techniques such as XGBoost, LightGBM, or deep neural networks, which may offer better performance and generalization for complex classification tasks. Ensemble methods or automated hyperparameter tuning could also be explored to further boost accuracy.

To increase the system’s scalability, it can be expanded to cover multiple states or nationwide deployments by integrating diverse datasets across regions. Additionally, embedding the solution into a cloud-based or mobile interface would enable on-the-spot eligibility checks by field officers or applicants themselves.

In the long term, the system can leverage emerging technologies such as edge computing for offline usage in remote areas, and incorporate explainable AI (XAI) techniques to ensure transparency and trust in decision-making.

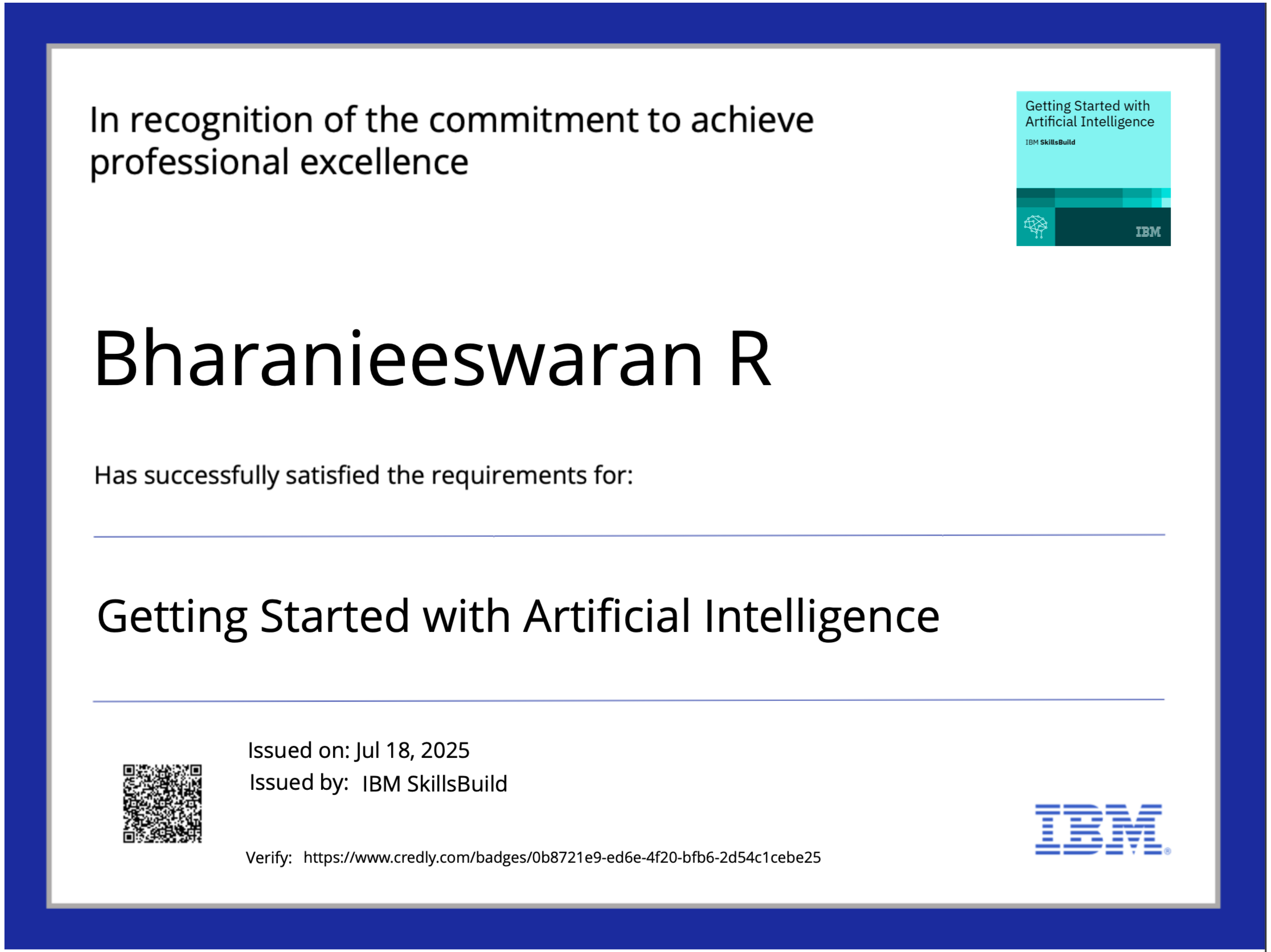
These future enhancements can help transform the current model into a robust, accessible, and intelligent eligibility recommendation engine that supports the equitable distribution of social benefits at scale.

References

- AI_KOSH Dataset – NSAP District-wise Pension Data
- https://aikosh.indiaai.gov.in/web/datasets/details/district_wise_pension_data_under_the_national_social_assistance_programme_nsap_1.html
- Scikit-learn: Machine Learning in Python
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011).
- Journal of Machine Learning Research, 12, 2825–2830.
- https://scikit-learn.org/stable/
- IBM Cloud Lite – Watson Studio Documentation
- https://www.ibm.com/cloud/watson-studio
- Understanding Random Forests: From Theory to Practice
- L. Breiman (2001), Random Forests, Machine Learning, Springer.
- https://link.springer.com/article/10.1023/A:1010933404324
- Best Practices for Data Preprocessing in Machine Learning
- Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer.
- ISBN: 978-3-319-14141-1
- Machine Learning Algorithms for Multi-class Classification
- Witten, I. H., Frank, E., & Hall, M. A. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
- ISBN: 978-0-12-804291-5

edunet foundation

IBM Certifications



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM SkillsBuild

IBM

Bharanieeswaran R

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/0b8721e9-ed6e-4f20-bfb6-2d54c1cebe25

IBM

IBM Certifications

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

Bharanieeswaran R

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

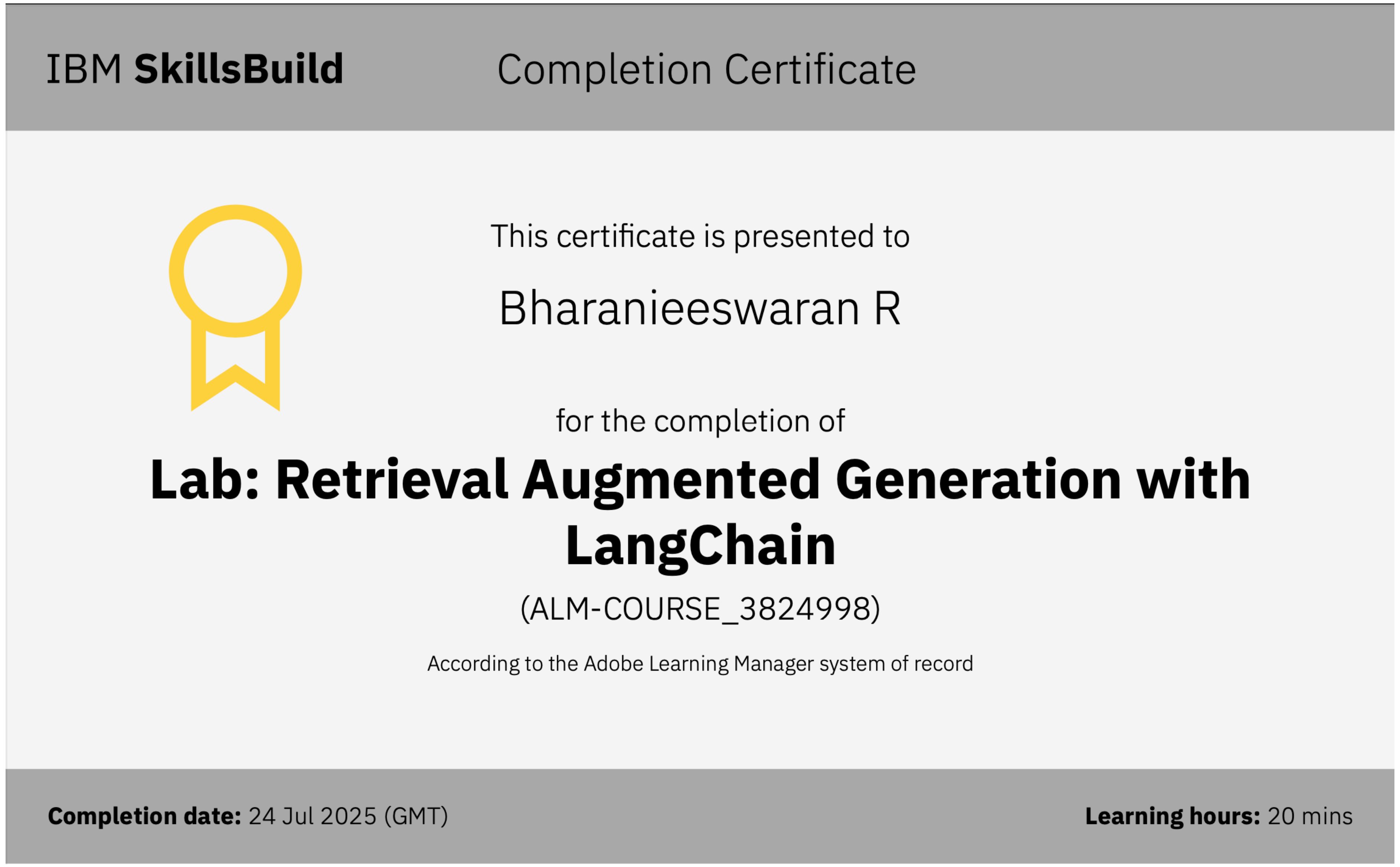
Issued on: Jul 18, 2025

Issued by: IBM SkillsBuild

IBM

Verify: https://www.credly.com/badges/f455d282-69fc-4032-9c91-7ba34243e145

IBM Certifications



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Bharanieeswaran R

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