
CAPSTONE PROJECT

NSAP ELIGIBILITY PREDICTION USING MACHINE LEARNING WITH IBM WATSONX.AI STUDIO

Presented By:

**Arul Murugan M - Chennai Institute Of Technology
- CSE**

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 welfare initiative by the Government of India to support elderly persons, widows, and individuals with disabilities living below the poverty line. Manually verifying applications and determining the correct pension scheme is time-consuming and often leads to delays or incorrect allocations.

These issues can prevent deserving citizens from receiving timely financial assistance, reducing the overall effectiveness of the program. As the number of applicants increases, it becomes difficult to manage the scheme eligibility process efficiently using manual methods.

PROPOSED SOLUTION

The proposed system aims to address the challenge of accurately determining the most suitable NSAP scheme for applicants based on their demographic and socio-economic characteristics. This involves leveraging machine learning techniques to automate and enhance the eligibility prediction process. The solution consists of the following components:

Data Collection:

- Gather historical applicant data from the **AI Kosh NSAP dataset**, which includes attributes like age, gender, disability status, BPL status, marital status, income, caste category, and location (state/district).

Data Preprocessing:

- Clean the dataset by handling missing values, removing duplicates, and managing inconsistencies.
- Normalize or scale numerical features (such as age and income) to improve model performance.
- Perform **feature selection** to retain only the most relevant variables affecting scheme eligibility.

Machine Learning Algorithm:

- Implement a **Random Forest Classifier**, a robust ensemble model that builds multiple decision trees and aggregates their predictions.
- Train the model on a labeled portion of the dataset (train-test split) to recognize patterns associated with each NSAP scheme.

PROPOSED SOLUTION

Deployment:

- Use **IBM Watsonx.ai Studio** on **IBM Cloud Lite** to train and deploy the model.
- Store the dataset in **IBM Cloud Object Storage** and use **AutoAI** or **Jupyter Notebooks** for modeling.
- Deploy the final trained model as a **Web Service Endpoint**, allowing real-time predictions based on new applicant data.

Evaluation:

- Evaluate model performance using multi-class metrics **Accuracy**.
- Test the model using multiple input scenarios to validate prediction consistency.
- Continuously monitor and improve model performance based on user feedback and new data.

Result :

- The **Random Forest Classifier** achieved an overall **accuracy of 98.4%**, effectively predicting the correct NSAP scheme based on applicant demographic and socio-economic data.
- Multiple **real-time predictions** were successfully performed using the deployed model in **IBM Watsonx.ai Studio**, and the outputs were verified using sample inputs and corresponding result images.

SYSTEM APPROACH

The "**System Approach**" section outlines the overall strategy and methodology followed in developing and deploying the NSAP scheme eligibility prediction system. This structured approach ensured the project was accurate, efficient, and deployable using IBM Watsonx.ai services.

System Requirements:

- **Platform:** IBM Cloud (Lite Plan)
- **Tool Used:** Watsonx.ai Studio
- **Data Format:** CSV file from AI Kosh NSAP dataset containing overall applicant details from various states.
- **Internet:** Required for cloud access and deployment
- **User Access:** IBM Cloud account and Watson Studio access

Tools and Platform Used:

- **IBM Watsonx.ai Studio:** Used to develop the ML model through visual pipeline (no-code environment).
- **Random Forest Classifier:** Selected automatically by Watsonx AutoAI as the best-performing model.
- **AutoAI Service:** Handled automated preprocessing, model selection, training, and optimization.

ALGORITHM & DEPLOYMENT

Algorithm:

- The **Random Forest Classifier** algorithm was used to train the model.
- It is a supervised machine learning algorithm that works by constructing multiple decision trees and combining their results for improved accuracy.
- The algorithm was automatically selected by **IBM Watsonx.ai Studio's AutoAI** as the best performer after testing multiple models.

Deployment Approach:

- The model was developed using the **AutoAI tool** inside **IBM Watsonx.ai Studio**, which required **no coding**.
- The dataset (CSV format) was uploaded to **IBM Cloud Object Storage** and connected directly to the AutoAI pipeline.
- Once training was complete, the best model (Random Forest) was deployed as a **Web Service Endpoint** using IBM Watsonx.ai.
- Predictions were generated in real-time by uploading new input data (CSV files and test cases).

RESULT

- The model trained using the **Random Forest algorithm** accurately predicted the correct NSAP scheme for each applicant.
- It was deployed successfully using **IBM Watsonx.ai Studio**, and prediction results were tested with real applicant data.
- Sample inputs were given, and the model returned correct scheme predictions with high accuracy.

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	00:00:35
	2	Pipeline 3	○ Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:26
	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

RESULT

NSAP_Prediction_Model_Deployment ✓ 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)
1	2025-2026	35	TAMIL NADU	10	CHENNAI	833	446	502
2	2025-2026	11	TAMIL NADU	15	DINDIGUL	444	189	201
3	2025-2026	33	TAMIL NADU	568	CHENNAI	38131	13286	24844
4								
5								

3 rows, 15 columns

Predict

RESULT

Prediction results

Prediction type

Multiclass classification

Prediction percentage



■ IGNDPS

■ IGNOAPS

Confidence level distribution

Display format for prediction results

☒ Table view ☐ JSON view

☐ Show input data ⓘ

	Prediction	Confidence
1	IGNDPS	90%
2	IGNDPS	70%
3	IGNOAPS	100%
4		
5		
6		
7		
8		
9		
10		
11		

Download JSON file

CONCLUSION

The NSAP Eligibility Prediction system was successfully developed using a no-code approach on **IBM Watsonx.ai Studio**. The primary goal was to automate the classification of applicants under the correct NSAP scheme using their demographic and socio-economic information. The use of the **Random Forest Classifier** helped in building a robust and accurate model, selected automatically by IBM AutoAI based on performance metrics. The dataset used contained real-world records from the **AI Kosh NSAP pension data**, covering multiple states and applicant categories.

Through proper preprocessing and visual model training, the system was able to generate reliable predictions and significantly reduce the manual workload involved in scheme allocation. The project demonstrated the power of cloud-based AI tools like Watsonx.ai in making machine learning accessible without the need for traditional coding. The final model was deployed as a web service, allowing government officers to input new applicant data and instantly receive scheme recommendations. Overall, the project improved efficiency, ensured better delivery of welfare, and proved the real-world impact of AI in public service domains.

FUTURE SCOPE

- **Enhanced Feature Set**

Future versions of the system can include additional features such as education level, occupation type, and household size. These variables can help the model better understand the applicant's socio-economic status and improve prediction accuracy.

- **Mobile and Web Application Development**

A user-friendly mobile app or web-based platform can be developed to allow government officials and applicants to access the prediction system easily. This would enable field-level officers to use the tool in real-time during field visits or data collection drives.

- **Integration with Government Databases**

The model can be connected with official databases like Aadhaar, income certificates, or BPL registers to automate and verify applicant details in real time. This will reduce manual verification and improve system reliability.

- **Multi-language Support**

Adding regional language support to the interface will make the system accessible to applicants and staff in different parts of India, especially in rural and underserved areas. It will promote inclusivity and broader adoption.

REFERENCES

- **AI Kosh NSAP Dataset**

National Social Assistance Programme (NSAP) – District-wise Pension Data

<https://aikosh.indiaai.gov.in>

- **IBM Watsonx.ai Documentation**

IBM Cloud Docs – Watsonx.ai Studio

<https://www.ibm.com/docs/en/watsonx>

- Breiman, L. (2001). **Random Forests.**

Machine Learning Journal, 45(1), 5–32.

<https://link.springer.com/article/10.1023/A:1010933404324>

- **IBM AutoAI Overview**

<https://www.ibm.com/cloud/watson-studio/autoai>

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Arul Murugan M

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 20, 2025

Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/4df4f4a7-a9e5-4bf6-be06-9576e0d2a697>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Arul Murugan M

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/3fb6d953-6e07-46e6-a874-fd6a086ef932>



IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to

Arul Murugan M

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