IBM AICTE PROJECT

PREDICTING ELIGIBILITY FOR NSAP SCHEMES USING MACHINE LEARNING

Presented By:
Akash Kumar Singh-Ajay Kumar Garg Engineering College
Ghaziabad-Computer Science and Engineering(CSE)

edunet foundation

OUTLINE

Problem Statement (NSAP Scheme Prediction Using Machine Learning)

Proposed System/Solution

System Development Approach (IBM Services, Auto Ai)

Algorithm & Deployment

Result (Output Image)

Conclusion

Future Scope

References

PROBLEM STATEMENT

Predicting Eligibility for NSAP Schemes using Machine Learning

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.

PROPOSED SOLUTION

We propose a machine learning-based multi-class classification model to predict the most suitable NSAP sub-scheme for an applicant using their demographic and socio-economic attributes. The model will be trained on district-level data such as gender distribution, caste categories, Aadhaar linkage, and beneficiary counts. After preprocessing and feature engineering, classification algorithms like Random Forest or XGBoost will be used. The trained model will be deployed using IBM Cloud Lite services to assist government officials in automating and improving the accuracy of scheme allocation.

Data Collection:

- **Gathered historical data on NSAP scheme distribution across Indian districts, including state, district, financial year, gender-wise beneficiary count, caste distribution (SC, ST, OBC, General), and Aadhaar/mobile linkage.**
- **Utilized government-published data from AI-Kosh: District-wise NSAP Pension Data to ensure data authenticity and policy relevance.**

Data Preprocessing:

- **Performed data cleaning and transformation**, including handling missing values, encoding categorical variables, and normalizing numerical features to prepare the dataset for model training.
- **Engineered relevant features** such as total beneficiaries by caste and gender, and removed redundant or non-informative columns (e.g., state/district names) to improve model accuracy.

Machine Learning Algorithm:

- **Used Snap Random Forest Classifier** with enhancements such as **Hyperparameter Optimization (HPO-1 and HPO-2)** and **Feature Engineering (FE)** to improve model performance and generalization.
- **Achieved a high classification accuracy of 98.4 %,** demonstrating the model's effectiveness in correctly predicting the **NSAP scheme (scheme_code)** based on socio-demographic inputs.

Deployment:

- **Deployed the trained Random Forest classification model** on **IBM Watsonx.ai Studio**, enabling real-time prediction of NSAP schemes through a cloud-based interface.
- **Leveraged IBM Cloud Lite services** for scalable, secure, and accessible model hosting, allowing stakeholders to test effectively.

edunet foundation

PROPOSED SOLUTION CONT…

Evaluation:

- **Achieved an overall accuracy of 98.4%**, indicating that the model performs exceptionally well in correctly classifying the **NSAP scheme based on input features.**
- **Used a confusion matrix and classification report** to evaluate class-wise performance, ensuring balanced predictions across **all scheme categories (IGNOAPS, IGNWPS, IGNDPS).**
- **Applied cross-validation techniques** to verify model stability and prevent overfitting, **ensuring consistent performance** across different subsets of the data.
- **Monitored precision, recall, and F1-score**, particularly to assess **model reliability** in correctly identifying **minority class schemes.**

SYSTEM APPROACH

This section outlines the overall strategy and methodology used to develop and implement the NSAP scheme prediction system.

System Requirements

- **Operating System**: Windows 10 / Ubuntu
- **Processor**: Intel i5 or higher
- **RAM**: Minimum 8 GB
- **Internet**: Required for IBM Cloud access and Watsonx.ai Studio

Libraries/Tools Required

- **Pandas** – for data manipulation and preprocessing
- **NumPy** – for numerical operations
- **Snap ML (IBM)** – for high-performance Random Forest with HPO and FE
- **Matplotlib / Seaborn** – for data visualization

ALGORITHM & DEPLOYMENT

Algorithm Selection:

- We selected the **Snap Random Forest Classifier**, a high-performance ensemble-based machine learning algorithm, well-suited for multi-class classification problems. It was chosen for its robustness, ability to handle high-dimensional data, and its integration with **IBM Watsonx.ai Studio**, which supports hyperparameter optimization (HPO) and feature engineering (FE). The model delivered a strong predictive performance with an accuracy of **98.4%**.

Data Input:

The model uses district-level socio-economic and demographic features as inputs, including:

- Gender-wise counts (totalmale, totalfemale, totaltransgender)
- Caste-wise counts (totalsc, totalst, totalobc, totalgen)
- Aadhaar and mobile linkage (totalaadhaar, totalmobilenumber)
- totalbeneficiaries, and geographic identifiers like district and state codes

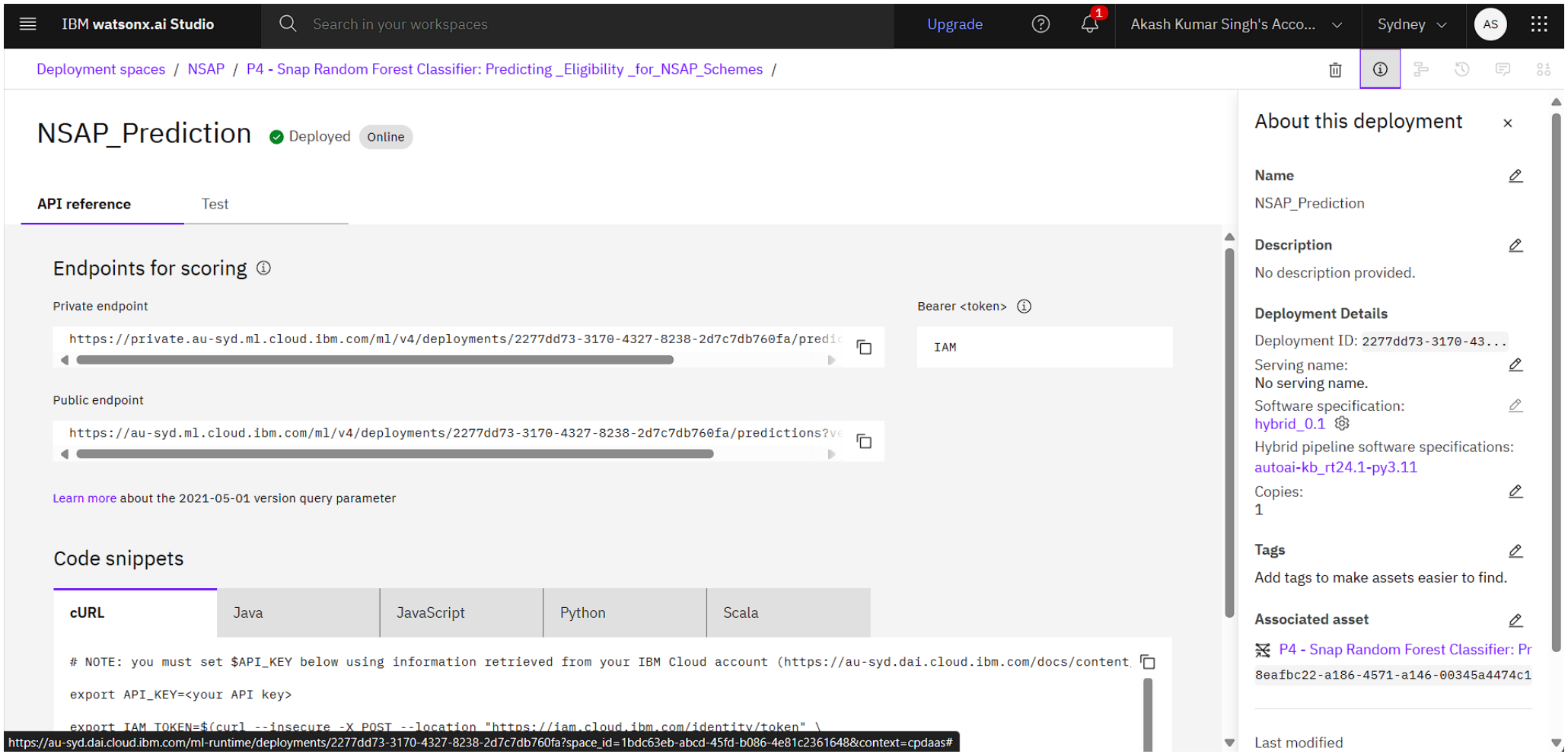
Training Process:

- The Random Forest model was trained using historical district-wise NSAP data. Enhancements such as **HPO-1 and HPO-2** were applied to fine-tune hyperparameters, and **Feature Engineering** was used to select and transform the most impactful attributes. Cross-validation was also used to validate the model's performance and prevent overfitting.

Prediction Process:

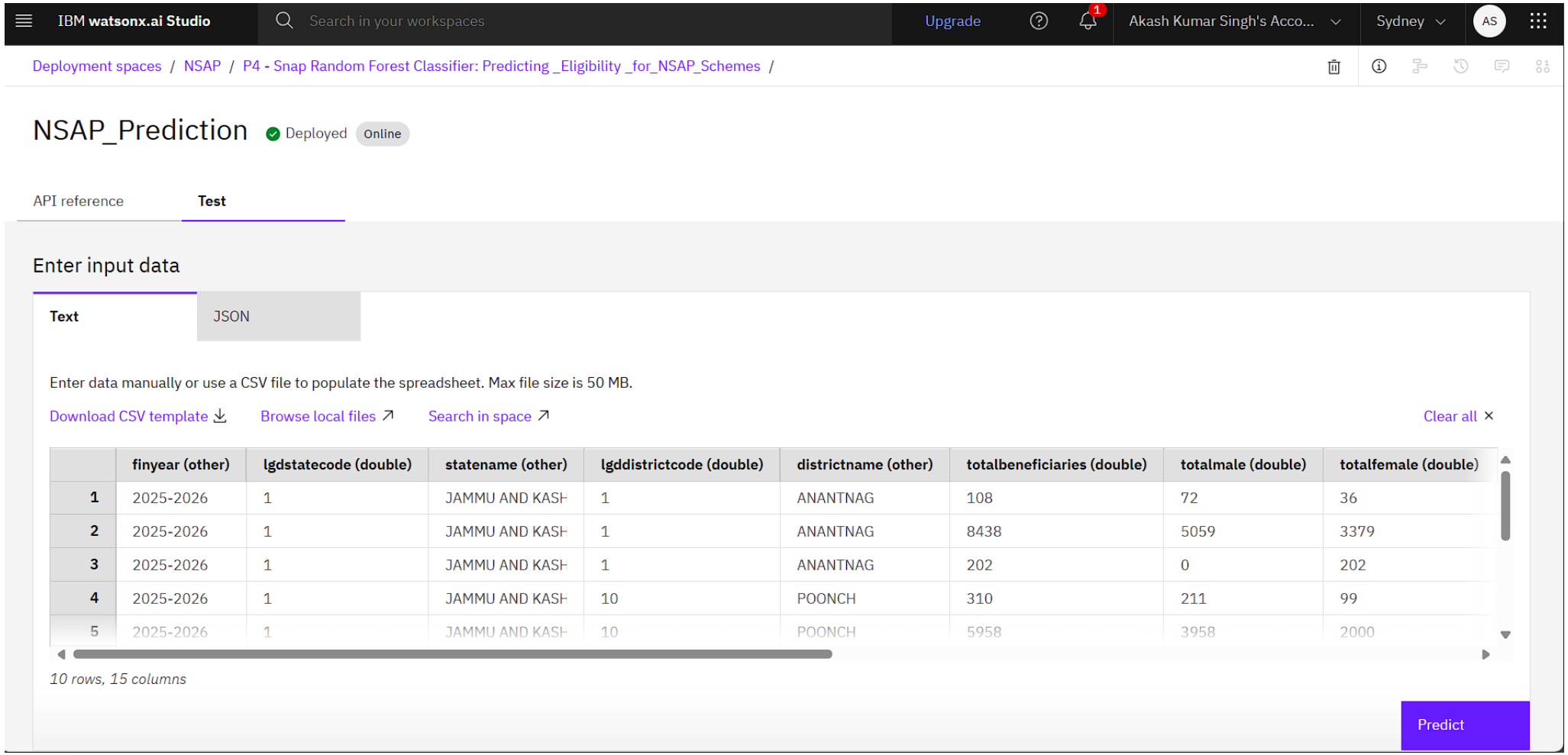
- Once trained, the model takes new socio-demographic data as input and predicts the most appropriate NSAP scheme (schemecode) an applicant or district may qualify for. The deployed model on IBM Watsonx.ai Studio allows for real-time or batch predictions, making it a scalable tool for government agencies to automate scheme classification.

RESULT



edunet foundation

RESULT CONT...



	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)
1	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	108	72	36
2	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	8438	5059	3379
3	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	202	0	202
4	2025-2026	1	JAMMU AND KASH	10	POONCH	310	211	99
5	2025-2026	1	JAMMU AND KASH	10	POONCH	5958	3958	2000

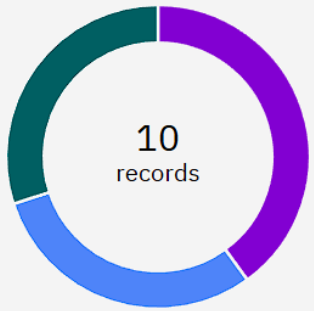
RESULT CONT…

Prediction results

Prediction type

Multiclass classification

Prediction percentage



IGNDPS IGNOAPS IGNWPS

Confidence level distribution

Display format for prediction results

(•) Table view ( ) JSON view

Show input data

	Prediction	Confidence	finyear	lgdstatecode	statename
1	IGNDPS	100%	2025-2026	1	JAMMU AN
2	IGNOAPS	70%	2025-2026	1	JAMMU AN
3	IGNWPS	100%	2025-2026	1	JAMMU AN
4	IGNDPS	100%	2025-2026	1	JAMMU AN
5	IGNOAPS	100%	2025-2026	1	JAMMU AN
6	IGNWPS	100%	2025-2026	1	JAMMU AN
7	IGNDPS	100%	2025-2026	1	JAMMU AN
8	IGNOAPS	80%	2025-2026	1	JAMMU AN
9	IGNWPS	100%	2025-2026	1	JAMMU AN
10	IGNDPS	100%	2025-2026	1	JAMMU AN

Download JSON file

edunet foundation

CONCLUSION

- Developed a multi-class classification model to predict the correct NSAP scheme for applicants.
- Used an enhanced Random Forest Classifier with feature engineering and hyperparameter optimization, achieving 98.4% accuracy.
- Automated scheme assignment reduces manual errors and speeds up application processing.
- Deployed the model on IBM Watsonx.ai, enabling secure and scalable access.
- The solution improves efficiency in welfare distribution and ensures benefits reach the right individuals faster.

GITHUB LINK

GitHub profile : Aakash4518 (Akash Kumar singh)

FUTURE SCOPE

The system can be further enhanced by incorporating more detailed applicant-level data to boost prediction accuracy. Integration with real-time government databases will enable dynamic, on-the-spot scheme assignment. A user-friendly mobile application can also be developed for citizens to check their eligibility instantly. Adding multilingual support will increase accessibility for diverse user groups across India. Additionally, explainable AI features can be integrated to provide transparency in decision-making, and fraud detection mechanisms can be included to identify suspicious or duplicate applications.

REFERENCES

- Data Source: District-wise Pension Data under NSAP
- Edunet Foundation – IBM SkillsBuild Program
- Cloud Services: IBM Watsonx.ai, IBM Cloud Object Storage
- Government Initiative Reference: National Social Assistance Programme (NSAP) – NSAP Official Site
- IBM Cloud Docs: https://cloud.ibm.com/docs

edunet foundation

IBM CERTIFICATIONS

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Akash Kumar Singh

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 16 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet foundation