CAPSTONE PROJECT

AUTOMATED NSAP SCHEME CLASSIFICATION USING MACHINE LEARNING FOR EFFICIENT SOCIAL WELFARE DISTRIBUTION

Presented By:
Name: Pranav Chauhan
Department: Computer Science & Applied Mathematics
College Name: Vivekananda Institute Of Professional Studies
University: Guru Gobind Singh Indraprastha University

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 automate the classification of applicants into the appropriate NSAP (National Social Assistance Program) sub-scheme, ensuring timely and accurate disbursement of social welfare benefits. This solution will use machine learning to analyse applicant data and recommend the most suitable scheme based on eligibility.**
- **Data Collection**
  - Gather historical applicant data, including demographic and socio-economic details such as gender, age group, caste category, Aadhaar and mobile availability, etc.
  - Include information about total beneficiaries and distribution across states and districts.
- **Data Preprocessing**
  - Clean and preprocess the dataset to handle missing, inconsistent, or duplicated entries.
  - Engineer relevant features from available fields (e.g., total SC/ST count, gender ratio, Aadhaar penetration) to enrich the model input.
  - Encode categorical features such as state/district names and normalize numerical columns.
- **Machine Learning Model**
  - Implement a multi-class classification algorithm (e.g., Random Forest, XGBoost, Logistic Regression) to predict the schemecode (e.g., IGNOAPS, IGNWPS, IGNDPS).
  - Perform hyperparameter tuning and cross-validation to optimize accuracy and generalization.
  - Evaluate feature importance to understand key factors affecting eligibility prediction.

- **Deployment**
  - Deploy the trained model using IBM Watsonx.ai and expose it via a REST API endpoint.
  - Create a lightweight interface or backend integration where government staff can submit new applicant data and receive immediate scheme predictions.
- **Evaluation & Monitoring**
  - Assess model performance using metrics like Accuracy, Precision, Recall, and F1 Score.
  - Continuously monitor predictions and collect feedback from domain experts to retrain and improve the model.
  - Ensure fairness and transparency in predictions to avoid bias in welfare distribution.
- **Result**

  A scalable, efficient, and intelligent classification tool that assists authorities in categorizing applicants under the correct NSAP scheme with high accuracy, reducing delays and manual errors in welfare distribution.

SYSTEM APPROACH

The **System Approach** section outlines the overall methodology, tools, and implementation strategy for predicting the appropriate NSAP (National Social Assistance Program) scheme using machine learning techniques.

- **System Requirements**
  - A machine with:
  - Minimum 4 GB RAM
  - Python 3.7 or above installed
  - Stable internet connection (for API access to IBM Cloud)
  - Access to IBM Watsonx.ai Studio (or Watson Machine Learning service)
  - GitHub (for version control and deployment logs)
- **Libraries Required**
  - Pandas
  - Numpy
  - Scikit-Learn
  - Matplotlib
  - Seaborn

ALGORITHM & DEPLOYMENT

- **This section outlines the machine learning algorithm used to classify applicants into the correct NSAP scheme, along with the deployment strategy used to make the model accessible and scalable.**
- **Algorithm Selection:**

A **multi-class classification algorithm** was selected for this problem, as the goal is to predict one of several NSAP schemes (e.g., IGNOAPS, IGNWPS, IGNDPS) based on applicant data. After exploring several models, **Random Forest Classifier** was chosen due to its:

- High accuracy on categorical and numerical features
- Ability to handle feature importance
- Robustness against overfitting on structured data
- Other algorithms like Logistic Regression and Decision Trees were also tested, but Random Forest offered a better balance ofperformance and interpretability.

- **Data Input:**

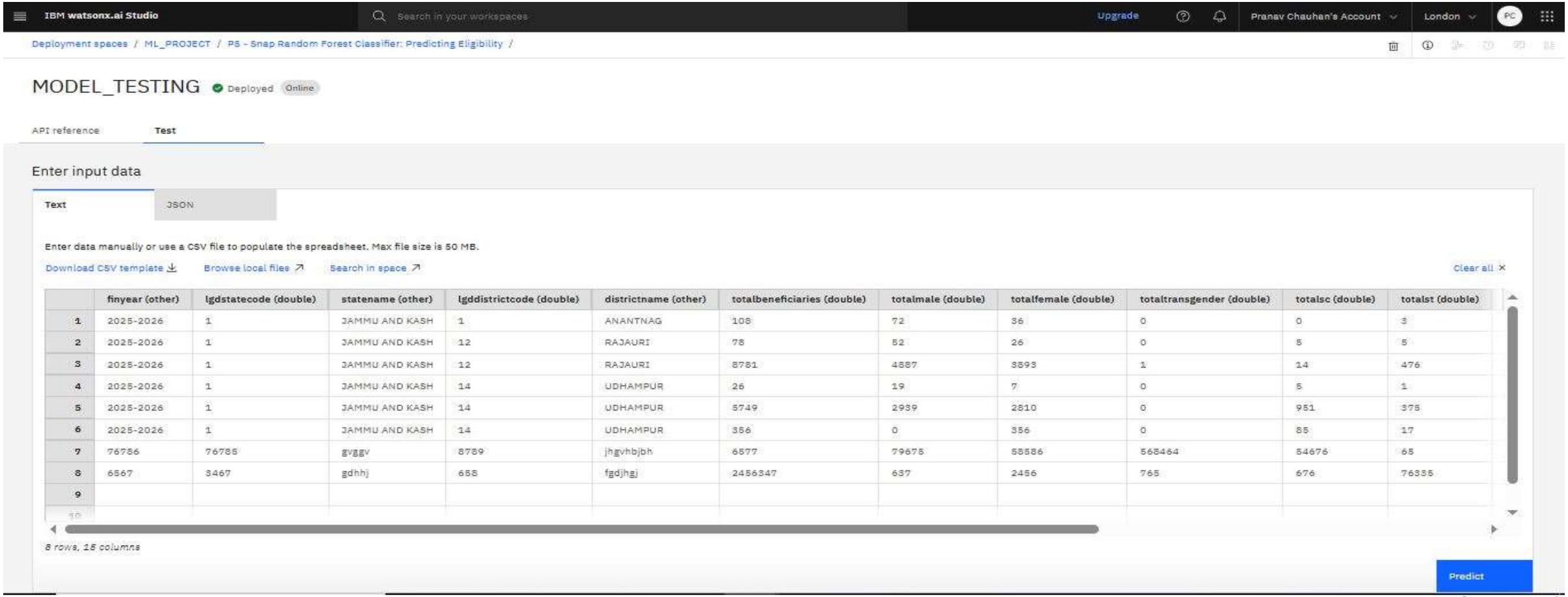
The model takes the following socio-economic and demographic features as input:

- totalmale ,totalfemale ,totaltransgender
- Caste Categories: totalgen ,totalobc ,totalsc ,totalst
- Accessibility Features: totalaadhaar ,totalmobilenumber
- Location Indicators: lgdstatecode ,lgddistrictcode

- **Training Process**
  - The model was trained on cleaned and preprocessed historical data, with the following steps:
  - Feature encoding for categorical variables (e.g., district names)
  - Train-test split (typically 80/20)
  - Model training using Random Forest with hyperparameter tuning (e.g., n_estimators ,max_depth)
  - Cross-validation to ensure generalization
  - Evaluation using Accuracy, Precision, Recall, and F1 Score to validate effectiveness
- **Prediction Process**
  - Once trained, the model is used to predict the schemecode for new input data. The prediction follows these steps:
  - Input data is formatted into the expected schema
  - A REST API request is made to the deployed model endpoint
- **Deployment Strategy**
  - The final model is deployed using **IBM Watson Machine Learning (WML)** via **Watsonx.ai Studio**:
  - The model is registered and deployed as an online REST API
  - API security is managed using Bearer Tokens from **IBM Cloud IAM**
  - External applications or scripts (e.g., in **Python**) can send **JSON** requests and receive real-time predictions
  - The deployment is scalable, reliable, and easy to monitor via the **IBM Cloud** dashboard

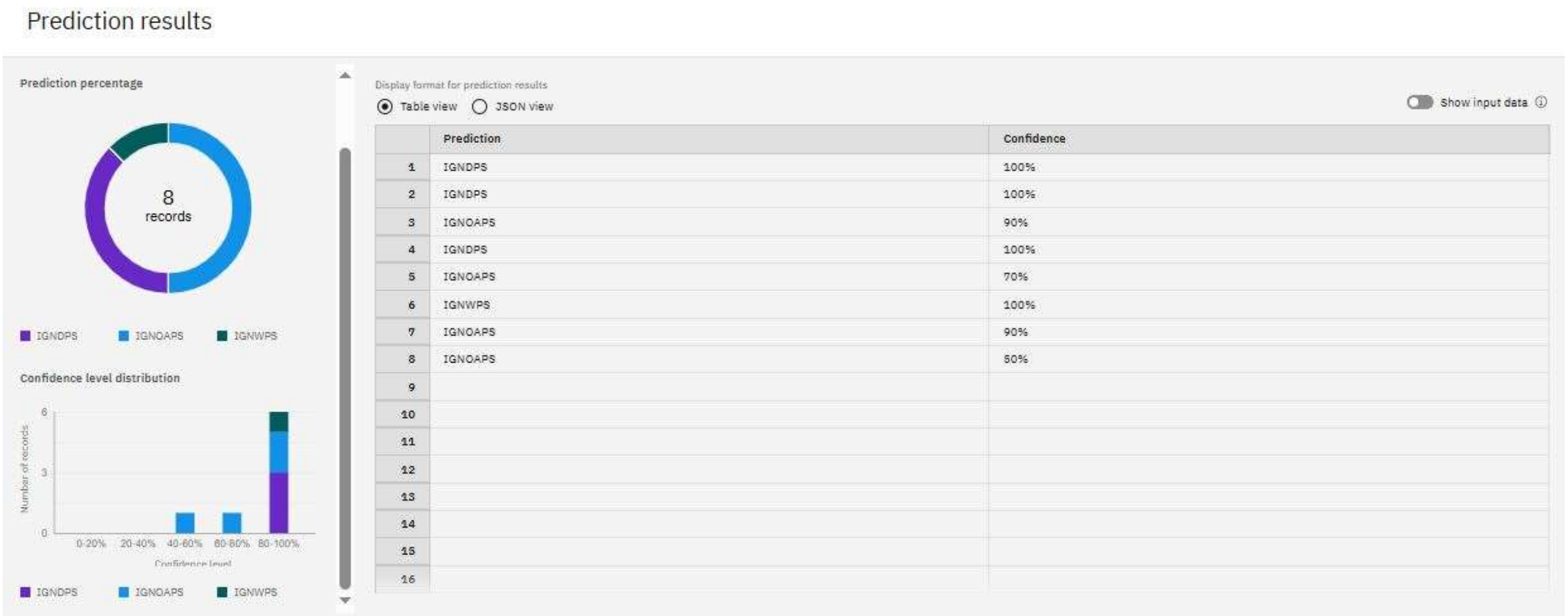
RESULT

- **TESTING INTERFACE OF A DEPLOYED RANDOM FOREST MODEL IN IBM WATSONX.AI STUDIO THAT PREDICTS THE SUITABLE NSAP SCHEME BASED ON APPLICANT DATA.**



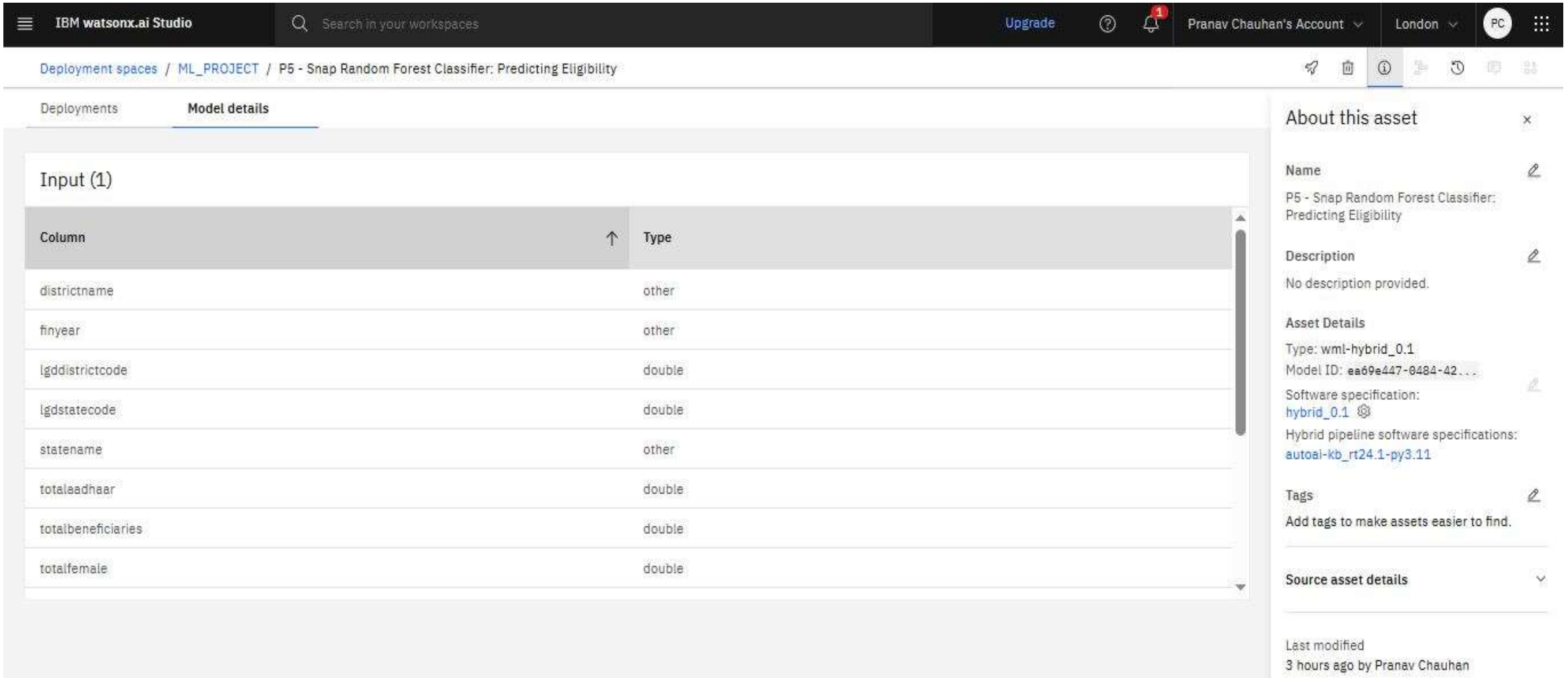
	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)	totaltransgender (double)	totalsc (double)	totalst (double)
1	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	108	72	36	0	0	3
2	2025-2026	1	JAMMU AND KASH	12	RAJAURI	78	52	26	0	5	5
3	2025-2026	1	JAMMU AND KASH	12	RAJAURI	8781	4887	3893	1	14	476
4	2025-2026	1	JAMMU AND KASH	14	UDHAMPUR	26	19	7	0	5	1
5	2025-2026	1	JAMMU AND KASH	14	UDHAMPUR	5749	2939	2810	0	951	375
6	2025-2026	1	JAMMU AND KASH	14	UDHAMPUR	356	0	356	0	85	17
7	76786	76786	gvggv	8789	jhgvhbjbh	6577	79676	58586	568464	54676	65
8	6567	3467	gdhhj	658	fgdjhgj	2456347	637	2456	765	676	76335
9											
10											

- **THIS IMAGE DISPLAYS THE MODEL'S PREDICTION RESULTS, SHOWING NSAP SCHEME CLASSIFICATIONS WITH THEIR RESPECTIVE CONFIDENCE LEVELS. MOST PREDICTIONS HAVE HIGH CONFIDENCE, INDICATING RELIABLE MODEL PERFORMANCE ACROSS THE TEST DATA.**



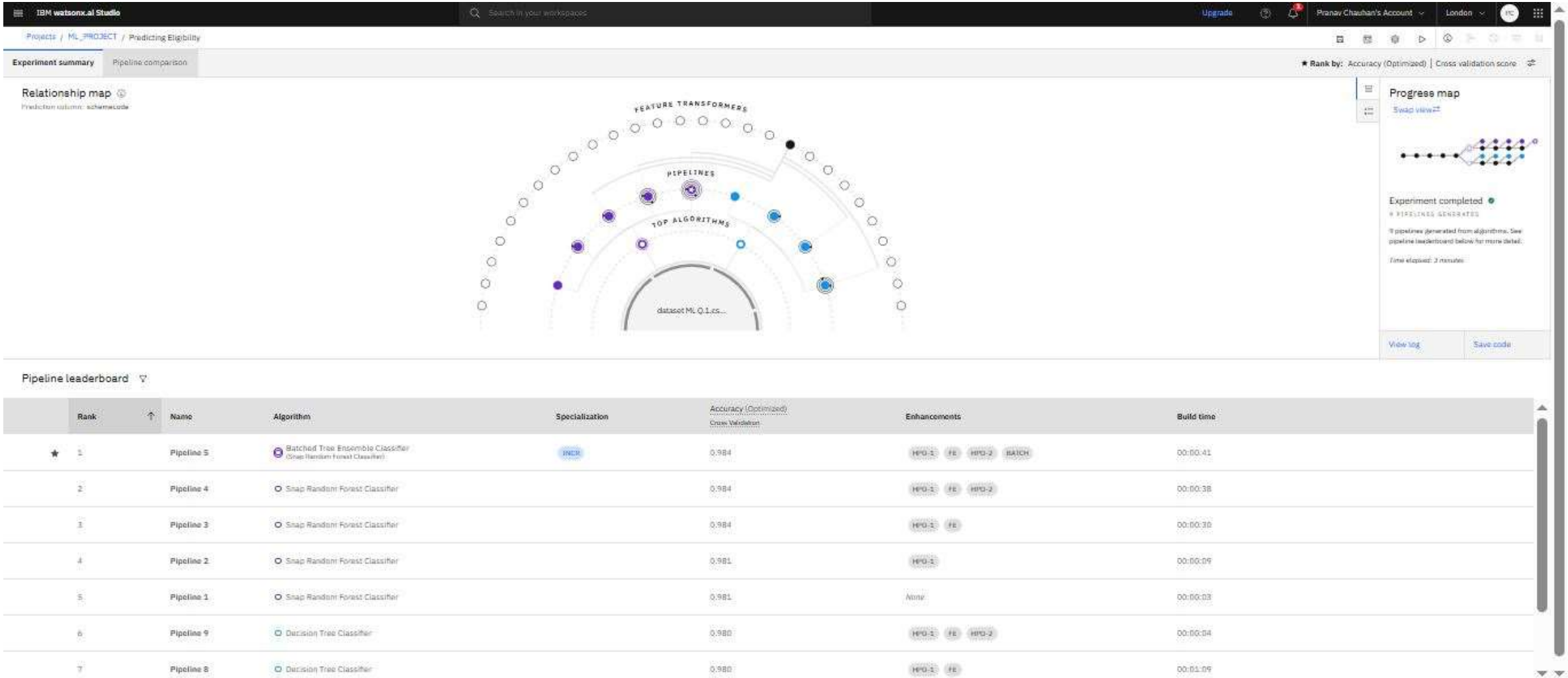
	Prediction	Confidence
1	IGNDPS	100%
2	IGNDPS	100%
3	IGNOAPS	90%
4	IGNDPS	100%
5	IGNOAPS	70%
6	IGNWPS	100%
7	IGNOAPS	90%
8	IGNOAPS	60%
9		
10		
11		
12		
13		
14		
15		
16		

- **MODEL TAKES INPUTS LIKE DISTRICT NAME, STATE CODE, TOTAL BENEFICIARIES, AND AADHAAR COUNT TO PREDICT NSAP SCHEME ELIGIBILITY USING A RANDOM FOREST CLASSIFIER IN IBM WATSONX.AI STUDIO.**



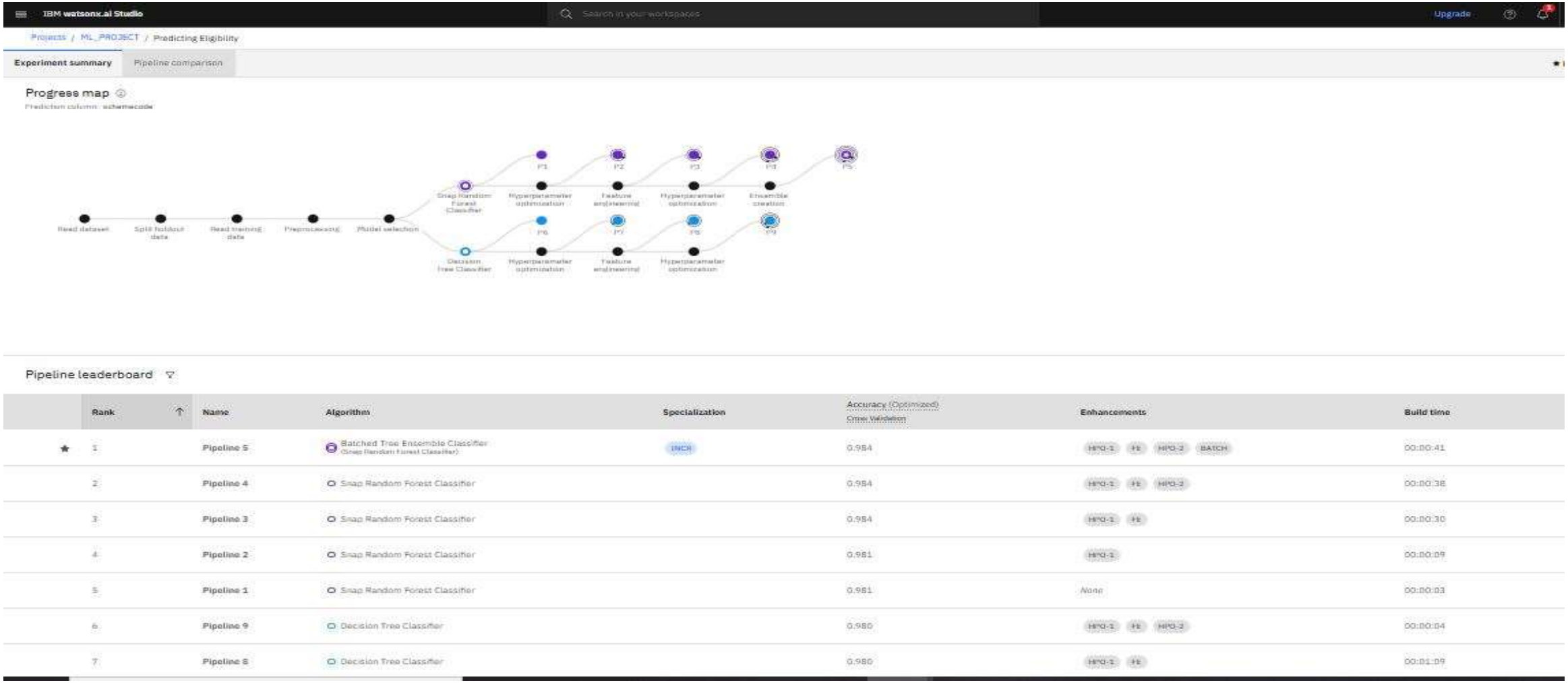
edunet foundation

- **AUTOAI RANKED MULTIPLE ML MODELS BASED ON ACCURACY. THE BEST MODEL WAS A BATCHED TREE ENSEMBLE CLASSIFIER WITH 98.4% ACCURACY, FOLLOWED BY MULTIPLE SNAP RANDOM FOREST CLASSIFIERS. EXPERIMENT COMPLETED IN UNDER 5 MINUTES WITH 9 PIPELINES GENERATED.**



edunet foundation

TOP PIPELINE (ENSEMBLE CLASSIFIER) ACHIEVED 98.4% ACCURACY USING AUTOAI WITH OPTIMIZATIONS LIKE HYPERPARAMETER TUNING AND FEATURE ENGINEERING. BUILT IN UNDER A MINUTE IN A DIFFERENT VIEW WHICH SHOECASING THE WHOLE MACHINE LEARNING PROCESS.



	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 5	Batched Tree Ensemble Classifier (Snap Random Forest Classifier)	INCR	0.984	HPO-1 FE HPO-2 BATCH	00:00:41
	2	Pipeline 4	Snap Random Forest Classifier		0.984	HPO-1 FE HPO-2	00:00:38
	3	Pipeline 3	Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:30
	4	Pipeline 2	Snap Random Forest Classifier		0.981	HPO-1	00:00:09
	5	Pipeline 1	Snap Random Forest Classifier		0.981	None	00:00:03
	6	Pipeline 9	Decision Tree Classifier		0.980	HPO-1 FE HPO-2	00:00:04
	7	Pipeline 8	Decision Tree Classifier		0.980	HPO-1 FE	00:01:09

edunet foundation

CONCLUSION

- The project addresses a key challenge: **Identifying the most suitable National Social Assistance Programme (NSAP) scheme for applicants based on demographic and socio-economic data**. Manual eligibility verification often leads to inefficiencies and misallocations.
- To solve this, we developed a machine learning-based prediction system using IBM Watsonx.ai. The model—powered by optimized ensemble techniques like Random Forest and Decision Trees—achieved high accuracy (98.4%) through automated pipeline generation, feature engineering, and hyperparameter tuning.
- The system has been deployed with an interactive interface, allowing users to input applicant data and receive real-time eligibility predictions. This ensures faster, fairer, and data-driven decision-making in welfare scheme distribution.
- **MY Github Repository Link:** https://github.com/Pranav-0922?tab=repositories

FUTURE SCOPE

- **Enhanced Data Inclusion**: Adding more detailed inputs like income level or disability status can help solve misclassification and increase prediction accuracy.
- **Advanced Algorithm Integration**: Using models like XGBoost or neural networks can further improve performance on complex data.
- **Geographical Expansion**: Scaling the system across cities and states can help reach more beneficiaries and ensure broader scheme coverage.
- **Edge Computing Deployment**: Running the model on local devices can solve connectivity issues in rural areas by enabling offline predictions.
- **Explainable AI (XAI)**: Introducing transparency in predictions can build trust and help officials understand eligibility decisions.
- **Continuous Learning via Feedback**: Adding a feedback loop will help the model learn from misclassifications and improve over time.
- **Multi-Scheme Recommendation**: Expanding the system to recommend multiple schemes based on user data can ensure no applicant is left out of rightful benefits.

REFERENCES

- **National Social Assistance Programme (NSAP)** – Government of India Welfare Scheme Guidelines
  https://nsap.nic.in/
- **AI Kosh Dataset** – Open government dataset used for training the model
  https://data.gov.in
- **IBM Watsonx.ai Studio** – Used for building, training, and deploying the ML model
  https://www.ibm.com/cloud/watsonx-ai
- **IBM Cloud** – Cloud infrastructure platform for hosting and managing AI projects
  https://www.ibm.com/cloud
- **IBM Cloud Assistant** – Tool used for setting up deployment and testing interfaces
  https://www.ibm.com/docs/en/cloud-paks/cp-data/4.7.x?topic=services-watson-assistant
- **Scikit-learn Documentation** – Python library for model building and evaluation
  https://scikit-learn.org/stable/
- **Towards Data Science** – Best practices for multi-class classification and model tuning
  https://towardsdatascience.com

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Pranav Chauhan

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/4ee8bc97-3276-4fac-aa6a-f3c67f1df154

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Pranav Chauhan

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