IBM AICTE PROJECT

PREDICTING ELIGIBILITY FOR NSAP WELFARE SCHEMES USING MACHINE LEARNING

Presented By:
Name : Khushi Basra
University : Graphic Era Deemed to be University, Dehradun
Department : Computer Science & Engineering

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References
- Github Link
- IBM Certifications

edunet foundation

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 address the challenge of accurately predicting the appropriate NSAP (National Social Assistance Programme) scheme for each applicant, based on their demographic and socio-economic information. By leveraging machine learning through IBM Watson's AutoAI, the solution will assist government bodies in automating the classification of applicants into the correct welfare scheme, ensuring timely and accurate benefit distribution. The solution will consist of the following components:

- Data Collection:
  - Utilize the dataset from AI Kosh containing district-wise records under the NSAP scheme.
  - Ensure the dataset includes relevant features such as gender composition (totalmale, totalfemale, totaltransgender), caste distribution (totalsc, totalst, totalobc, totalgen), Aadhaar and mobile availability (totalaadhaar, totalmobilenumber), and geographic information (statename, districtname, lgdstatecode, lgddistrictcode).
- Data Preprocessing:
  - Clean the dataset to handle missing or inconsistent values.
  - Define the target variable as the scheme code (schemecode), which represents different NSAP schemes (e.g., IGNOAPS, IGNWPS, IGNDPS).
- Machine Learning Model via AutoAI:
  - Use IBM Watson AutoAI to automatically build, optimize, and evaluate multiple machine learning pipelines for multi-class classification.
  - AutoAI will perform: Data preprocessing, Model selection (e.g., Random Forest, Logistic Regression, Gradient Boosting), Hyperparameter optimization, Performance comparison across models.

PROPOSED SOLUTION

- Deployment:
  - Deploy the best-performing model as an online REST API using IBM Watson Machine Learning.
  - The API will accept applicant data as input and return the predicted scheme name.
  - This enables integration with government portals or mobile applications for real-time classification.
- Evaluation:
  - Evaluate the model using appropriate classification metrics such as : Accuracy, Precision, Recall, and F1-Score, Confusion Matrix.
  - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.
- Result:
  - The resulting model will serve as a decision-support tool for government agencies to:
  - Automatically categorize applicants under the correct NSAP sub-scheme.
  - Reduce manual verification workload.
  - Minimize allocation errors and delays.
  - Improve the overall efficiency and transparency of welfare scheme distribution.

SYSTEM APPROACH

The system approach outlines the technical methodology adopted to design and implement a machine learning-based model that accurately classifies applicants under the appropriate NSAP welfare scheme. The solution leverages IBM Cloud and Watson Studio's AutoAI capabilities to automate the model development, evaluation, and deployment processes:

- System requirements:
  - As the entire machine learning workflow is executed using IBM Cloud's managed services, the local system requirements are minimal. A standard computer with a stable internet connection and a modern web browser, such as Google Chrome or Microsoft Edge, is sufficient to interact with IBM Watson Studio. No high-end processing or storage capabilities are needed locally, as computation, data handling, and training take place entirely on the cloud.
  - To build and deploy the solution, the following IBM Cloud services are used:
    - IBM Watson Studio for creating the project, configuring the AutoAI experiment, and managing assets such as datasets and models.
    - IBM Cloud Object Storage to store and retrieve the dataset securely within the IBM Cloud project.
    - IBM Watson Machine Learning to deploy the trained model as a REST API endpoint that can be accessed in real time for predictions.

SYSTEM APPROACH

- **Libraries Required to Build the Model:**
  - While IBM AutoAI handles most of the machine learning workflow automatically within the cloud environment, underlying technologies rely on several well-established Python libraries and frameworks. These include:
    - Pandas and NumPy for internal data preprocessing and manipulation.
    - Scikit-learn (sklearn) for model building, evaluation, and classification metrics.
    - XGBoost and LightGBM for gradient boosting models, which are frequently selected by AutoAI for their high accuracy on structured datasets.
    - Matplotlib and Seaborn, used during internal model visualization.
    - IBM Watson Machine Learning SDK (ibm-watson-machine-learning) when interacting with the deployed model via API in an external application or script.

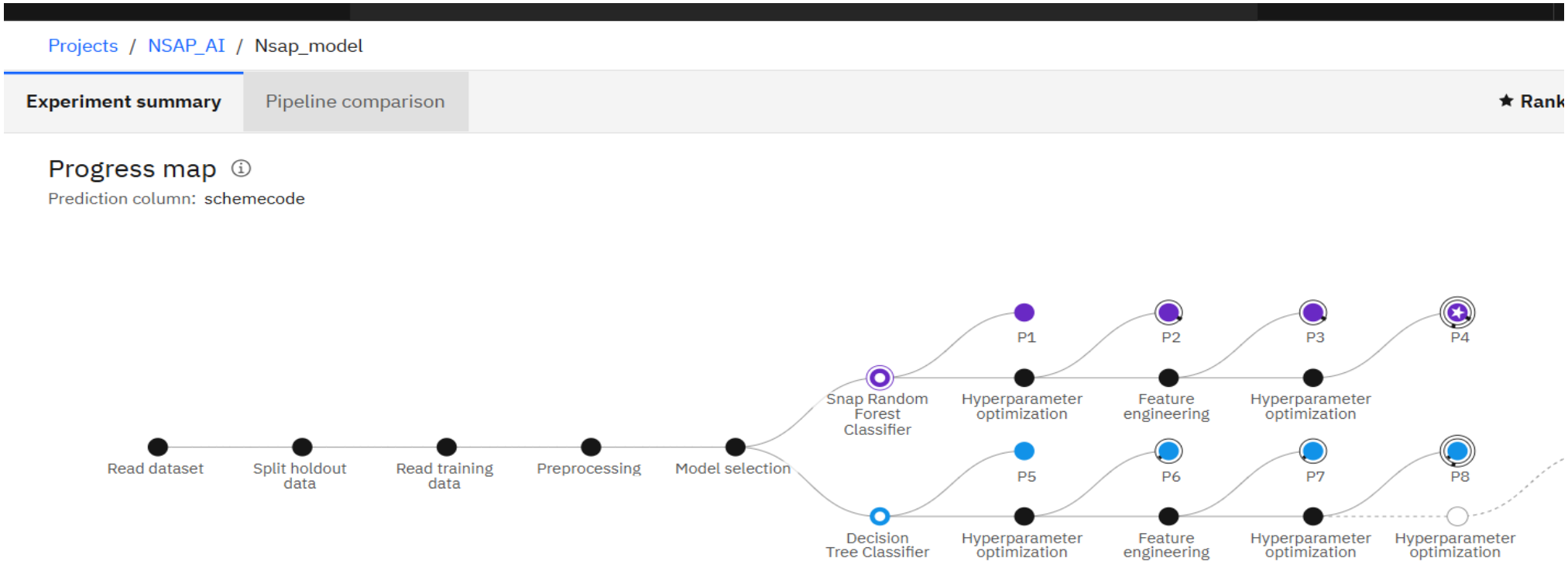
ALGORITHM & DEPLOYMENT

- This section outlines the algorithm selection and training methodology for developing the NSAP scheme classification model using IBM AutoAI.
- Algorithm Selection:
  - To automate the classification of applicants into appropriate NSAP schemes, the problem was framed as a multi-class classification task. IBM Watson AutoAI was used to automatically identify the best-performing algorithms for this type of problem. AutoAI evaluated several machine learning models, including: Logistic Regression (Multinomial), Random Forest Classifier, Gradient Boosting (XGBoost/LightGBM), Decision Trees.
  - These algorithms are well-suited for categorical outputs with multiple classes (IGNOAPS, IGNWPS, IGNDPS).
  - They handle both numerical and categorical features effectively.
  - Ensemble methods (e.g., Random Forest, XGBoost) generally provide higher accuracy and robust performance in structured datasets.
  - The final model was selected based on evaluation metrics including Accuracy, Macro-Averaged F1-Score, and ROC AUC, as automatically ranked in AutoAI's leaderboard.
- Data Input:
  - The dataset consists of aggregated district-level features such as gender distribution, caste category distribution, and Aadhaar/mobile coverage. Predictions are made at the group level, not for individual applicants.
  - The target variable used for classification is schemecode, which represents the NSAP scheme assigned to that district.

ALGORITHM & DEPLOYMENT

- **Training Process:**
  - The training process was fully managed by IBM AutoAI, which involved:
    - Data Splitting: Automatic train-test split (e.g., 80%-20%).
    - Preprocessing: AutoAI handled encoding of categorical features and normalization of numerical data.
    - Model Selection: Multiple pipelines were trained using different classifiers.
    - Hyperparameter Optimization: AutoAI applied internal hyperparameter tuning to improve performance.
    - The top-performing pipeline was saved for deployment.
- **Prediction Process:**
  - Once trained, the model can predict the most appropriate NSAP scheme for a new applicant based on their input data.
  - Prediction Steps include :
    - Input includes aggregated district-level features such as totalmale, totalfemale, totalsc, totalst, totalobc, totalgen, totalaadhaar, totalmobilenumber, along with statename and districtname.
    - Data is preprocessed as per the training pipeline.
    - The model classifies the input into one of the NSAP schemes (e.g., IGNOAPS).
    - The result is returned via the deployed REST API on IBM Watson Machine Learning.
    - This process enables real-time or batch predictions, allowing integration into government systems for faster decision-making.

RESULT



AutoAI Pipeline: This flow diagram illustrates how IBM AutoAI automatically constructs and evaluates multiple model pipelines.

edunet foundation

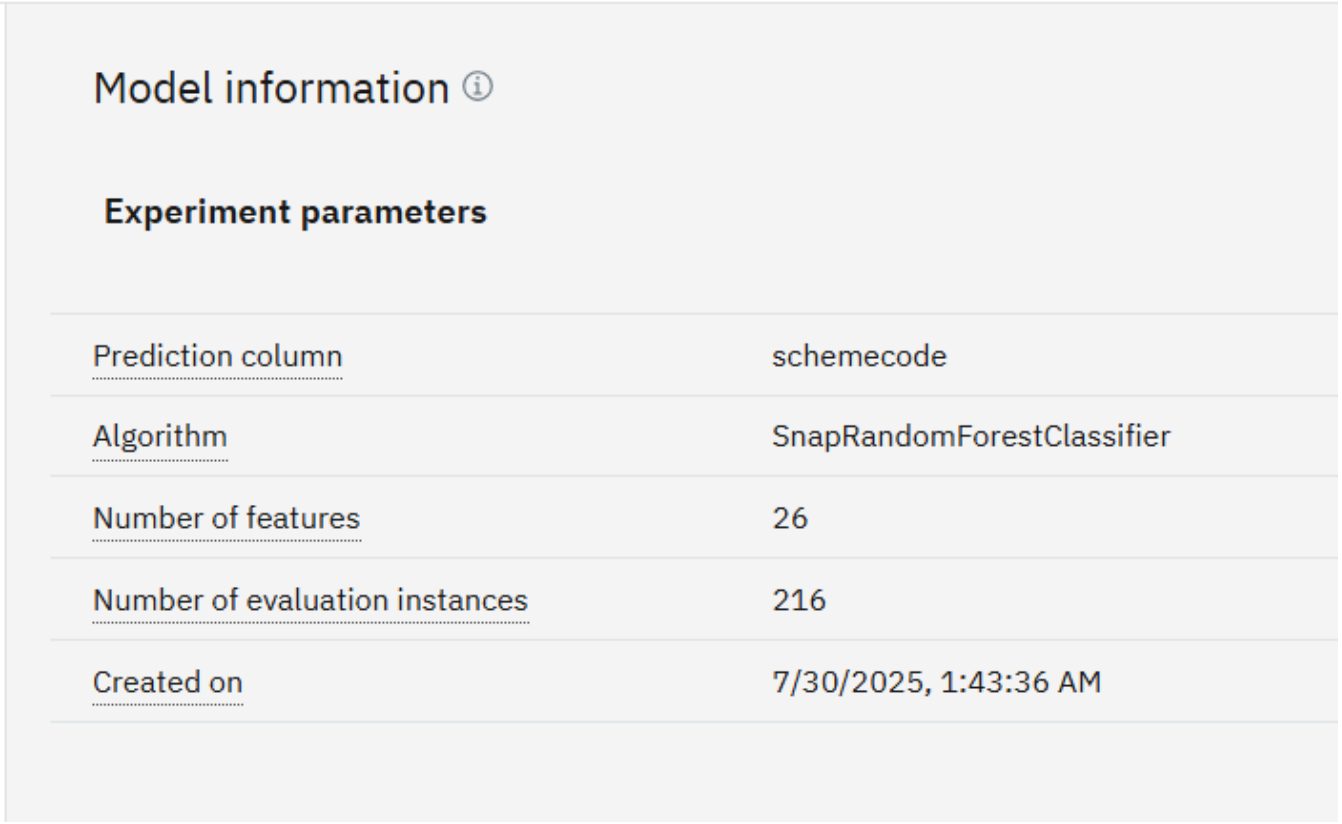
RESULT

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	00:00:37
	2	**Pipeline 3**	Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:29
	3	**Pipeline 2**	Snap Random Forest Classifier		0.981	HPO-1	00:00:09
	4	**Pipeline 1**	Snap Random Forest Classifier		0.981	*None*	00:00:03

Pipeline Leaderboard: Leaderboard displaying top 4 model pipelines ranked by F1-Score and Accuracy. The best model was selected from here.

RESULT



Model information

Experiment parameters

Prediction column	schemecode
Algorithm	SnapRandomForestClassifier
Number of features	26
Number of evaluation instances	216
Created on	7/30/2025, 1:43:36 AM

Model Information: Summary of the top-performing model pipeline selected by AutoAI. This model demonstrated the highest accuracy and F1-score among all evaluated pipelines.

RESULT

Model evaluation measure

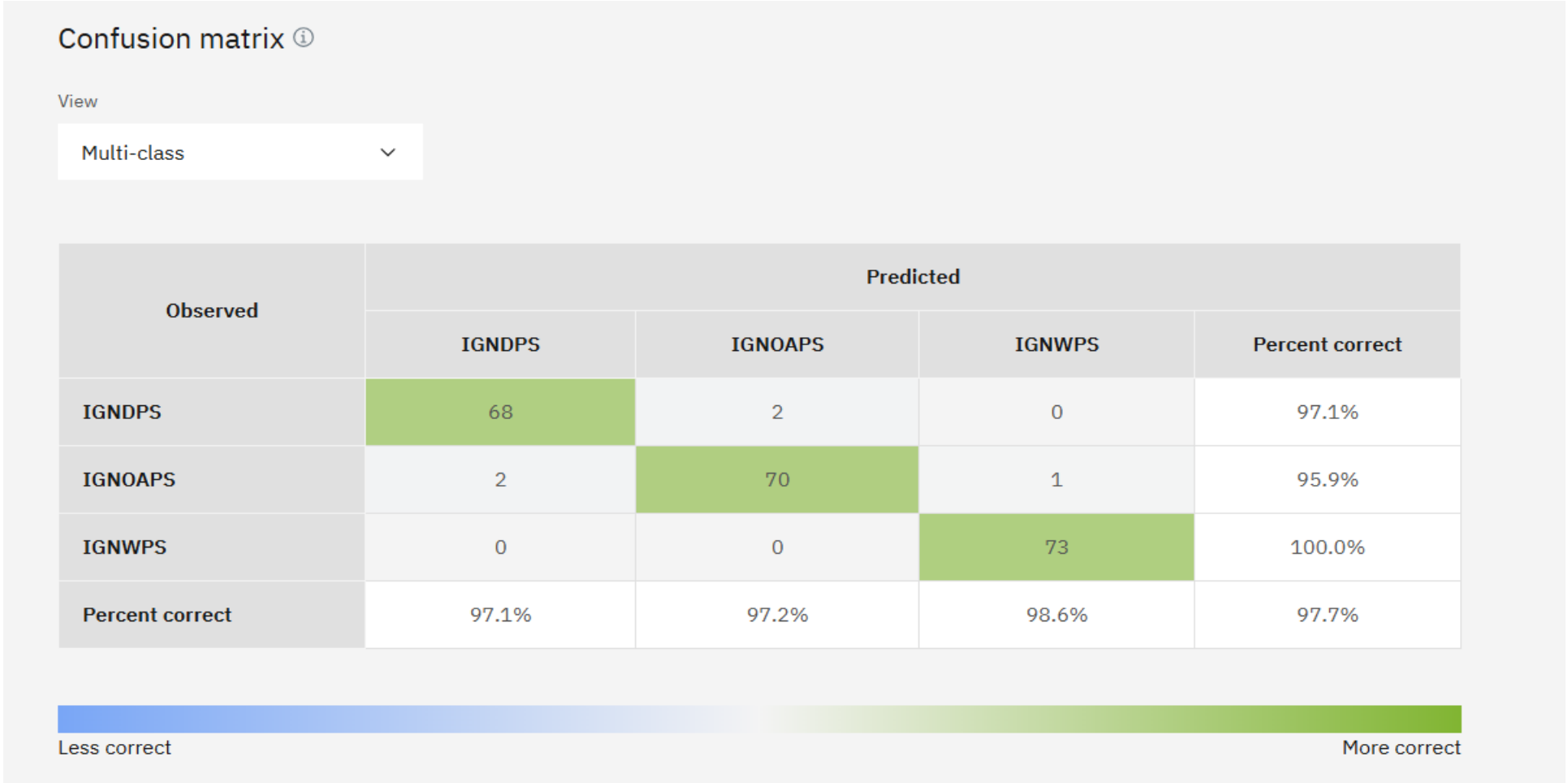
View

Multi-class

Measures	Holdout score	Cross validation score
Precision macro	0.977	0.984
Accuracy	0.977	0.984
Recall macro	0.977	0.984
Weighted precision	0.977	0.984
F1 macro	0.977	0.984
Weighted f1 measure	0.977	0.984
Weighted recall	0.977	0.984
Log loss	0.222	0.143

Evaluation Metrics: Performance metrics of the selected model, including Accuracy, Precision, Recall, and F1-Score. These metrics confirm the model's effectiveness in classifying NSAP scheme eligibility.

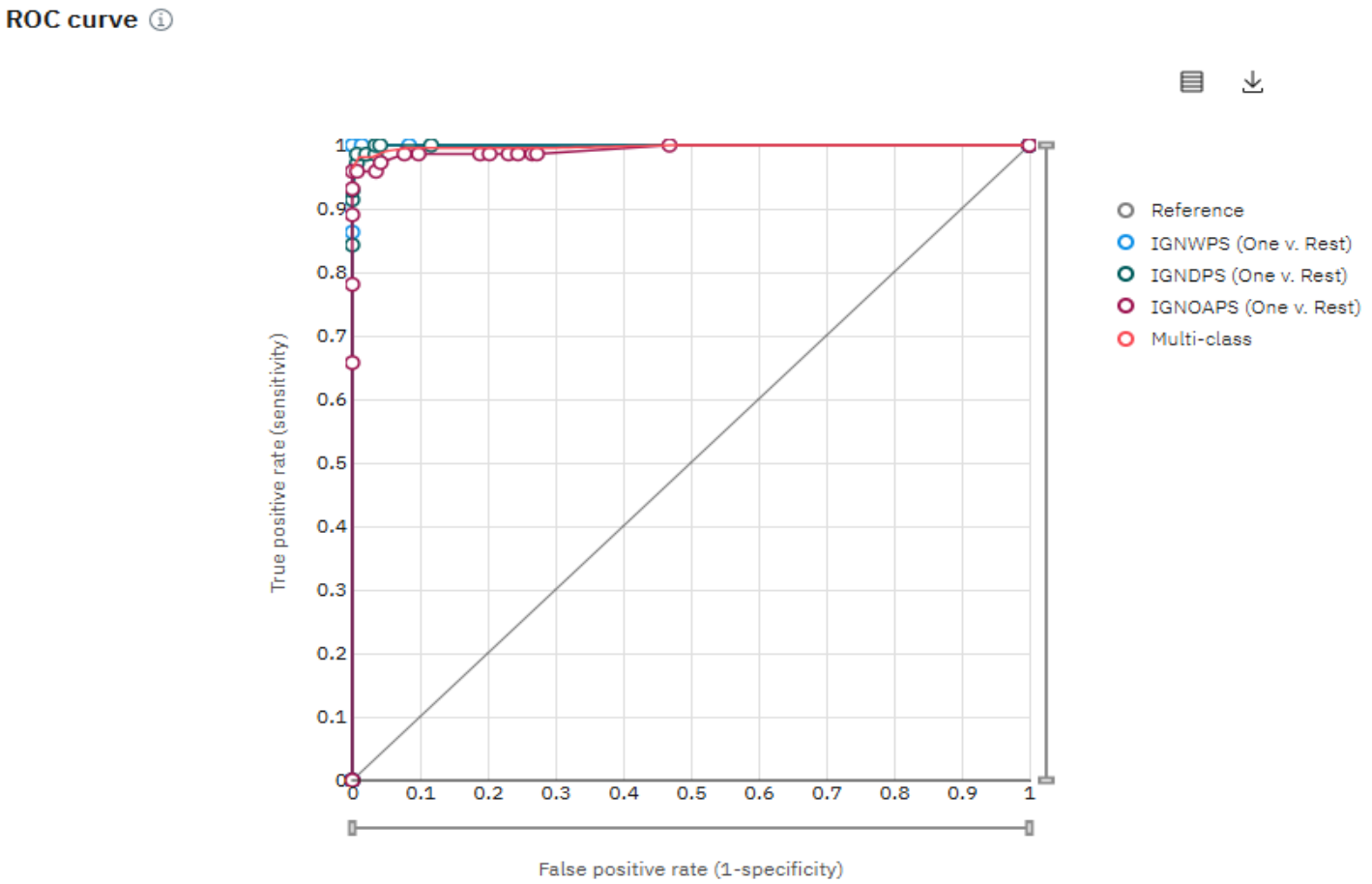
RESULT



Confusion Matrix: Shows actual vs predicted class distributions, helping assess classification performance.

edunet foundation

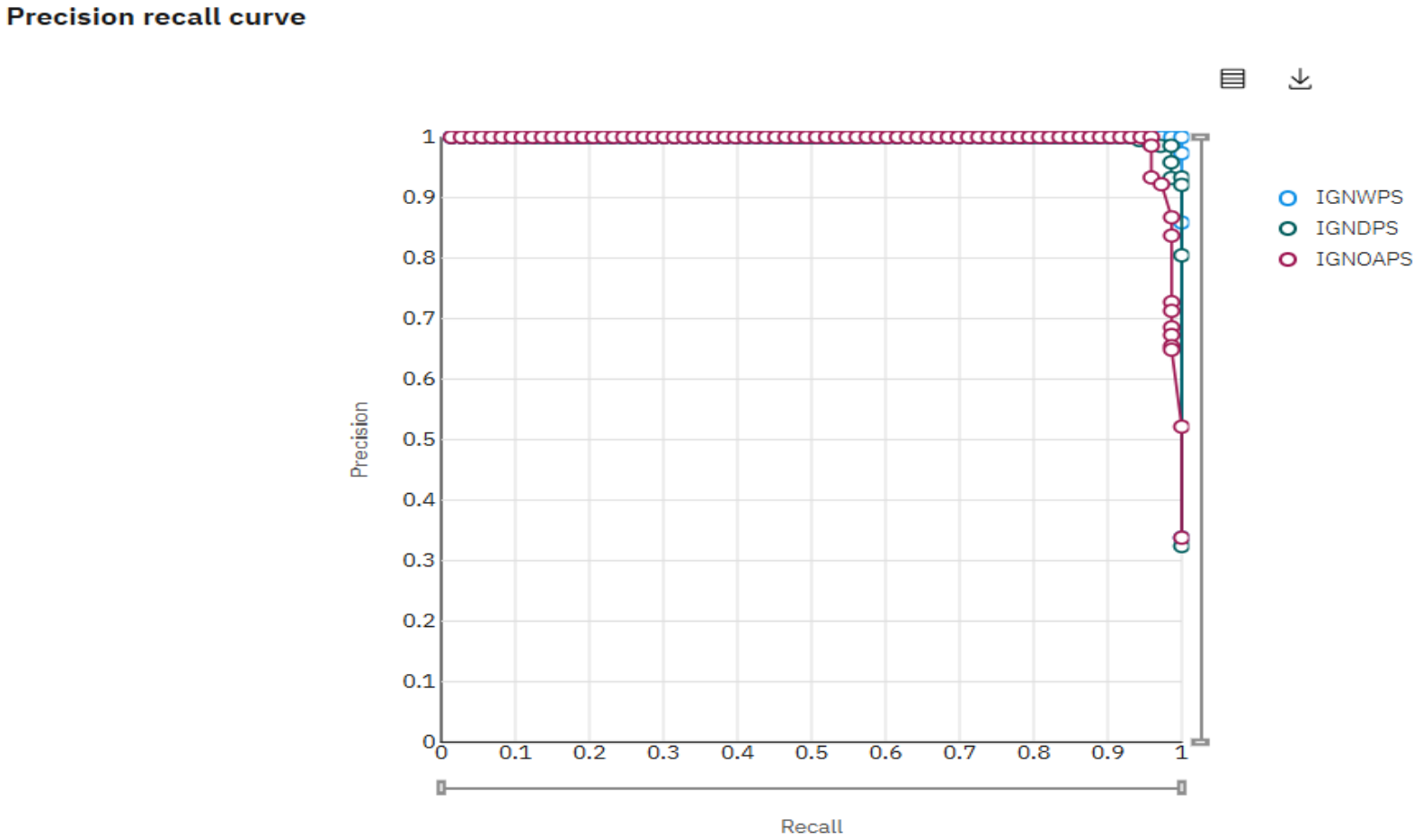
RESULT



ROC Curve: Receiver Operating Characteristic (ROC) curve showing the trade-off between true positive rate and false positive rate for each NSAP class. AUC scores close to 1 indicate strong class separation.

edunet foundation

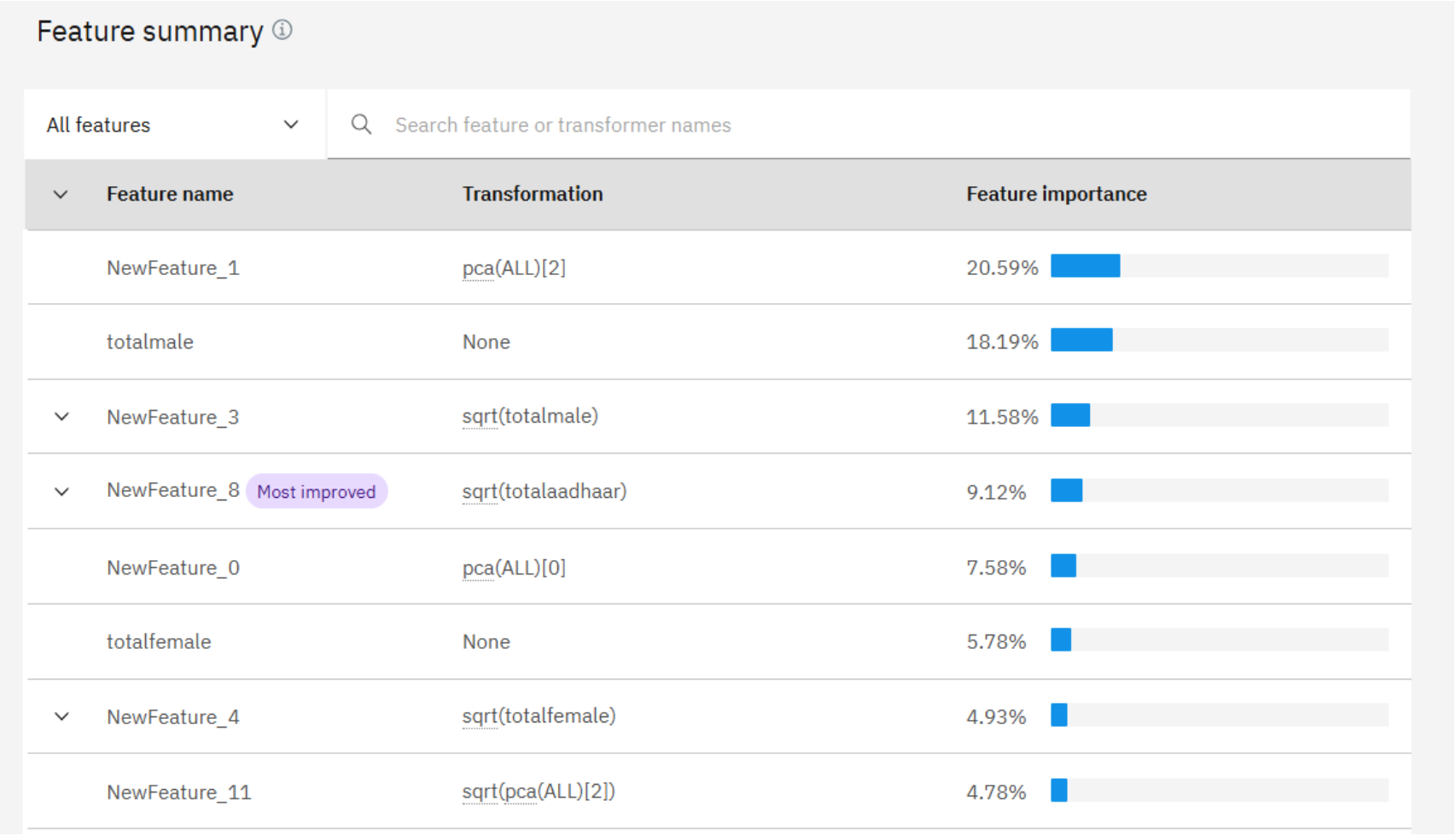
RESULT



Precision Recall Curve: Precision-Recall curve for each class label, helpful in evaluating model performance under class imbalance. High area under the curve signifies reliable predictions.

edunet foundation

RESULT



Feature summary

All features

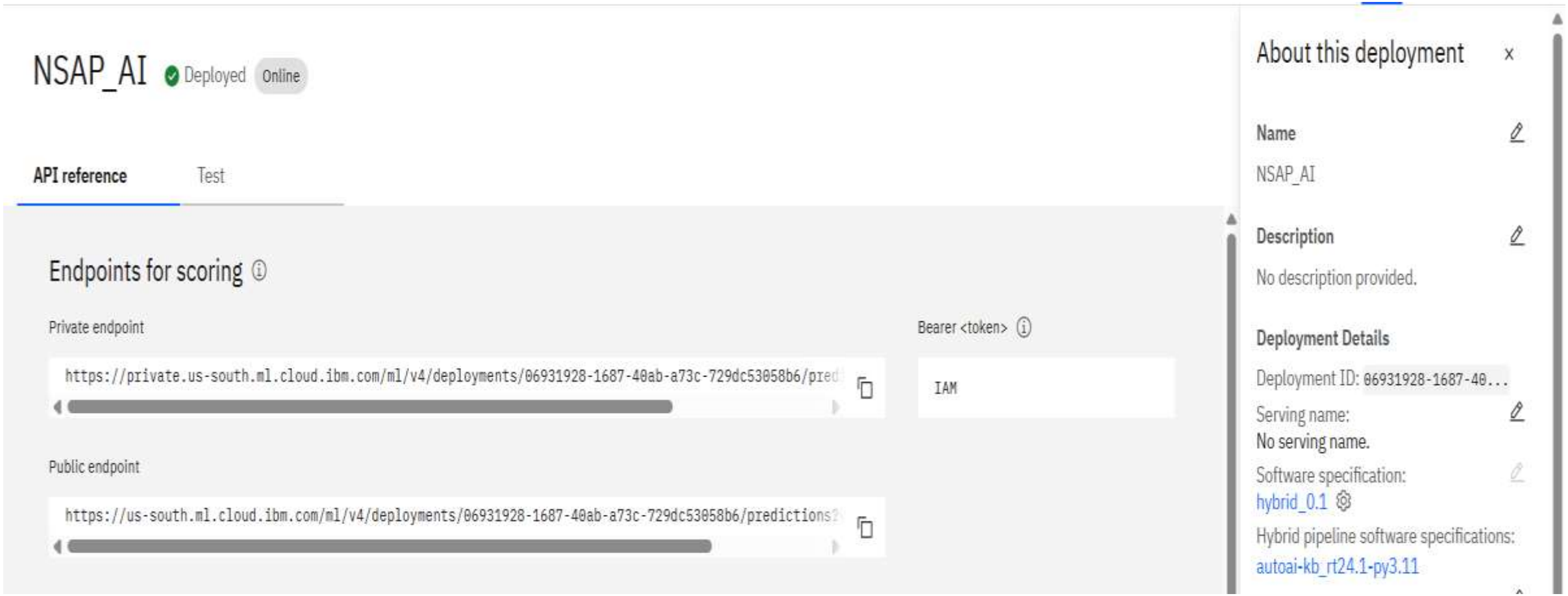
Search feature or transformer names

	Feature name	Transformation	Feature importance
	NewFeature_1	pca(ALL)[2]	20.59%
	totalmale	None	18.19%
⌄	NewFeature_3	sqrt(totalmale)	11.58%
⌄	NewFeature_8 Most improved	sqrt(totalaadhaar)	9.12%
	NewFeature_0	pca(ALL)[0]	7.58%
	totalfemale	None	5.78%
⌄	NewFeature_4	sqrt(totalfemale)	4.93%
	NewFeature_11	sqrt(pca(ALL)[2])	4.78%

Feature Importance: Ranked list of features based on their influence on model predictions.

edunet foundation

RESULT



Deployed Model : The trained model was deployed as a REST API using IBM Watson Machine Learning. The public endpoint allows external applications to send input data and receive NSAP scheme predictions in real time.

edunet foundation

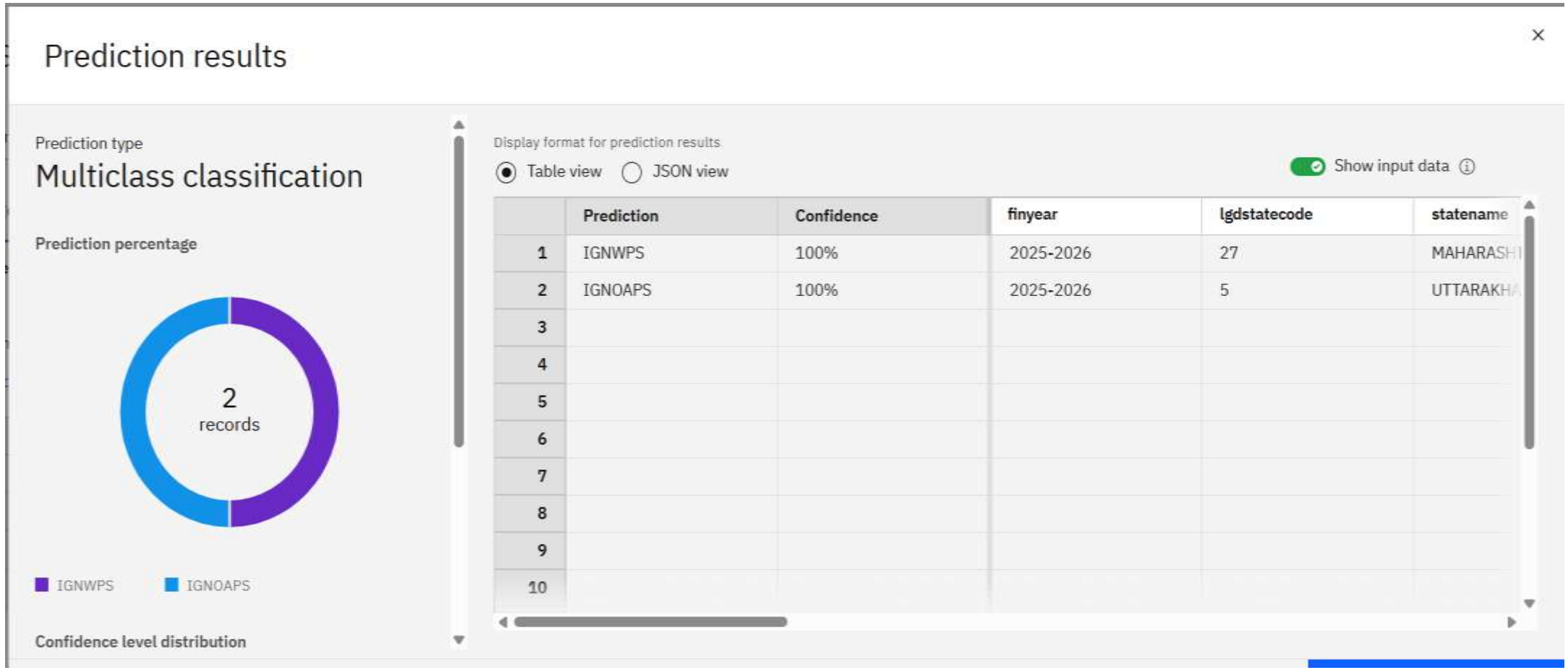
RESULT

	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)	totaltransgender (double)	totalsc (double)	totalst (double)
1	2025-2026	27	MAHARASHTRA	490	PUNE	782	0	782	0	57	22
2	2025-2026	5	UTTARAKHAND	49	DEHRADUN	27037	13700	13337	0	1815	7370

Input Data: Sample input tested against the deployed model using Watson Machine Learning's built-in test interface.

edunet foundation

RESULT



Prediction Results: Sample prediction results generated by the deployed AutoAI model. This output demonstrates how the model returns an NSAP scheme based on the provided input features.

edunet foundation

CONCLUSION

- This project successfully leverages IBM AutoAI to develop a machine learning-based eligibility prediction system for the National Social Assistance Programme (NSAP). AutoAI streamlined the entire model lifecycle—from data preprocessing and algorithm selection to hyperparameter optimization and evaluation—significantly reducing manual effort while ensuring high model performance.
- The system effectively analyzes key demographic and socio-economic indicators to predict the most appropriate NSAP scheme with speed and accuracy. While the current dataset is at the district level and captures aggregate information such as gender distribution, caste category breakdown, and Aadhaar/mobile availability, the model still offers meaningful insights into regional eligibility trends. By automating this process, the system helps reduce manual workload, minimize human bias, and improve transparency and consistency in welfare scheme allocation.
- During the implementation, challenges such as missing values, class imbalance, and interpretability were encountered. IBM AutoAI addressed many of these automatically through techniques like automated imputation, feature transformation, and internal hyperparameter tuning. The leaderboard of models, evaluation metrics, and feature importance charts generated by AutoAI provided actionable insights and guided the final model selection. The deployment of the top-performing model as a REST API using IBM Watson Machine Learning also enables real-time predictions, allowing seamless integration into government platforms and mobile applications.
- Overall, this AI-driven approach not only enhances the accuracy and efficiency of eligibility predictions but also sets the stage for scalable, data-driven governance. With further refinements—such as incorporating individual-level data, real-time updates, and multilingual support—this solution can evolve into a powerful decision-support tool across a broader range of welfare programs, ultimately contributing to a more inclusive and responsive public service ecosystem.

FUTURE SCOPE

- **Integration of Additional Data Sources:** Incorporate data from Aadhaar-linked government records, healthcare databases, census data, and financial history to improve prediction accuracy.
- **Edge Computing Deployment**: Deploy models on edge devices in remote service centers or local governance offices to enable faster and offline eligibility predictions in rural areas.
- **Multilingual User Interfaces:** Integrate the system with IBM Watson Assistant or similar tools to allow applicants to interact with the model using voice or text in their native language.
- **Continuous Learning and Model Updating:** Implement pipelines that allow the model to retrain itself periodically with newly collected data to remain updated with policy or demographic changes.
- **Integration with Government Portals:** Seamlessly integrate the system with official NSAP application platforms to automate decision-making and reduce manual intervention.
- **Enhanced Security and Compliance:** Apply secure data handling practices using IBM Cloud's identity, access management, and data encryption features to ensure privacy and compliance with government standards.

REFERENCES

- Ministry of Rural Development, Government of India. "National Social Assistance Programme (NSAP)" – https://nsap.nic.in
- IBM Documentation. "IBM watsonx.ai and AutoAI: Automating machine learning pipelines" – https://www.ibm.com/docs/en/watsonx
- Géron, A. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow", 2nd Edition, O'Reilly Media, 2019.
- IBM Cloud Docs. "Data Refinery and AutoAI for Automated ML Pipelines" – https://cloud.ibm.com/docs/autoai
- Scikit-learn developers. "User Guide: Supervised Learning Algorithms"–https://scikit-learn.org/stable/user_guide.html

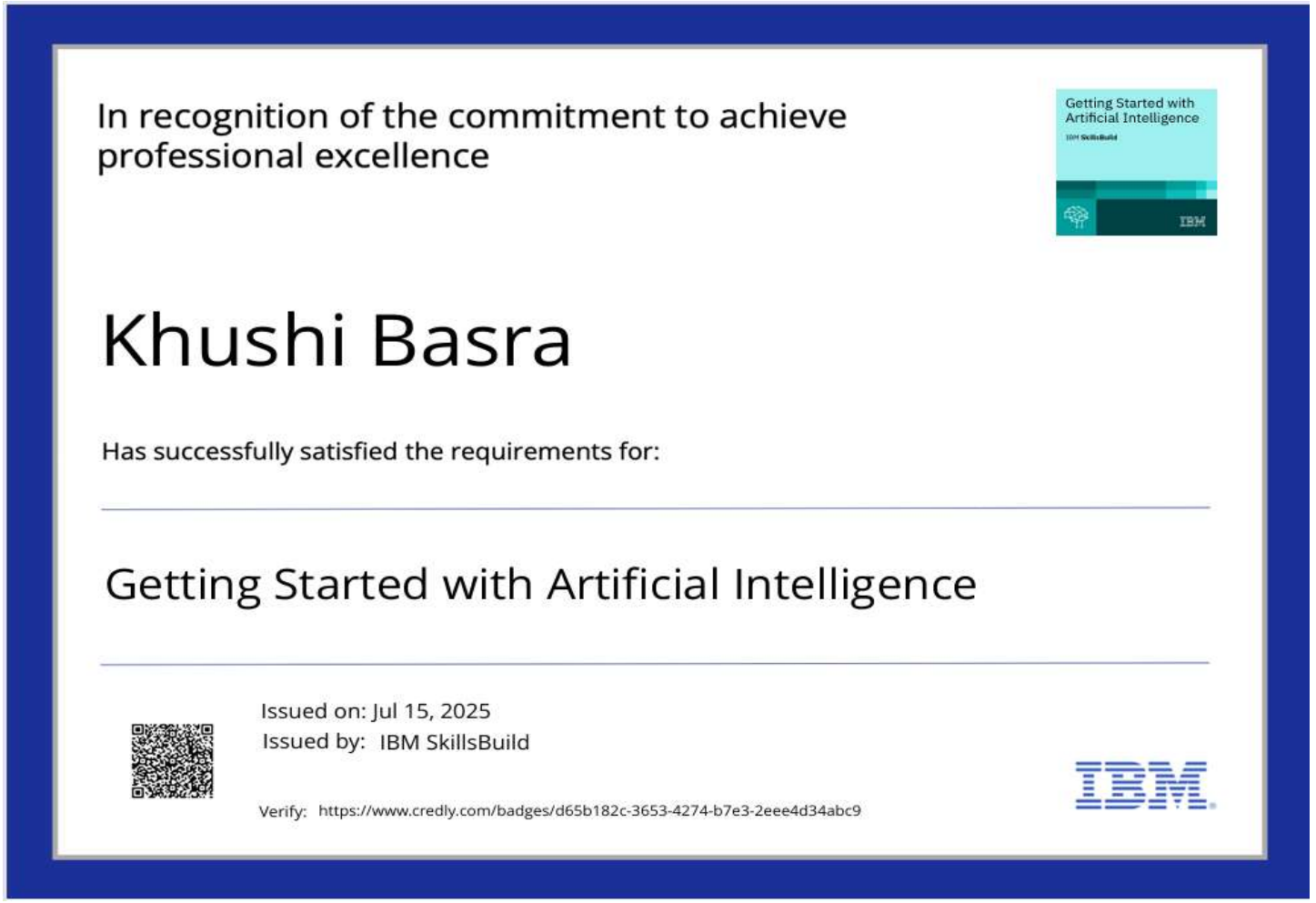
GITHUB LINK

- Github Link : https://github.com/khushibasra11/NSAP-Eligibility-Prediction

edunet foundation

IBM CERTIFICATIONS

- Credly certificate(getting started with AI)



edunet foundation

IBM CERTIFICATIONS

- Credly certificate(Journey to Cloud)

IBM CERTIFICATIONS

- Screenshot / credly certificate(RAG Lab)

edunet foundation

THANK YOU

edunet foundation