CAPSTONE PROJECT

ANALYZING DEMOGRAPHIC AND REGIONAL DISPARITIES IN TELE-LAW CASE REGISTRATIONS FOR INCLUSIVE LEGAL ACCESS

Presented By:

1. Student Name- Shwati Narayan
2. College Name- KCC Institute of Technology and Management
3. Department- Bachelor of Technology CSE(Data Science)

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

- Despite the widespread expansion of the Tele-Law initiative across Indian states and districts, disparities still exist in how different communities and regions access these legal services. Marginalized groups such as SC, ST, and OBC populations often remain underrepresented in case registrations. Additionally, uneven distribution of Common Service Centres (CSCs) and varying regional outreach efforts contribute to unequal access.
- This project aims to **analyze gender-wise, caste-wise, and region-wise Tele-Law case registration data** to uncover demographic utilization patterns and assess the inclusiveness of the legal aid system. A data-driven approach is essential to **evaluate equity**, identify under-served areas, and support strategies for **improving accessibility and service delivery**.

PROPOSED SOLUTION

- By leveraging data analytics and AI tools on IBM Cloud, the solution provides insights into gender-wise, caste-wise, and region-wise representation to ensure more inclusive legal access. The solution will consist of the following components:

- **Data Collection**
  - Utilize official dataset of district-wise Tele-Law case registrations and advice-enabled data from FY 2021–22 to 2024–25
  - Extract features such as gender, caste (SC/ST/OBC/General), district, number of CSCs, and total registrations

- **Data Preprocessing**
  - Clean and structure the dataset to remove duplicates, handle missing values, and correct inconsistencies
  - Apply encoding for categorical variables and normalization for numerical data

- **Machine Learning & Analysis**
  - Use IBM AutoAI to predict "Total Registrations" based on demographic and geographic features
  - Apply classification/regression models to detect usage disparities across regions and social groups

- **Visualization & Insights**
  - Generate dashboards to show trends in gender-wise and caste-wise participation
  - Identify districts with low outreach among marginalized communities

- **Deployment**
  - Host the predictive model and insights dashboard using IBM Cloud Lite services
  - Ensure accessibility through a user-friendly interface for stakeholders and policy makers

- **Evaluation**
  - Evaluate model performance using metrics like R^2 score and MAE
  - Continuously refine the model with updated data for improved accuracy

- **Result**

  An AI-powered system that delivers deep insights into inclusion gaps within the Tele-Law initiative, helping optimize outreach and ensure equitable legal aid access across all communities

edunet foundation

SYSTEM APPROACH

- This section outlines the strategic methodology used to analyze demographic and regional disparities in Tele-Law case registrations. The system leverages IBM Cloud Lite services and AutoAI to perform data-driven analysis for inclusive legal access.
- System Requirements
  - IBM Cloud Lite Account
  - IBM Watson Studio / AutoAI access
  - Cleaned CSV dataset with district-wise Tele-Law data
  - A stable internet connection
  - Minimum 8 GB RAM and Intel i5 processor or equivalent
  - Modern web browser (Chrome, Firefox, Edge)
- Libraries/Tools Required to Build the Model
  - **pandas** – for data cleaning and manipulation
  - **numpy** – for numerical operations
  - **matplotlib / seaborn** – for visual representation of disparities
  - **scikit-learn** – for optional ML model experiments
  - **IBM AutoAI** – for automated model selection and training
  - **Watson Studio Visualization** – for generating dashboards and reports
  - **MS Excel / OpenRefine** – for basic preprocessing if needed

ALGORITHM & DEPLOYMENT

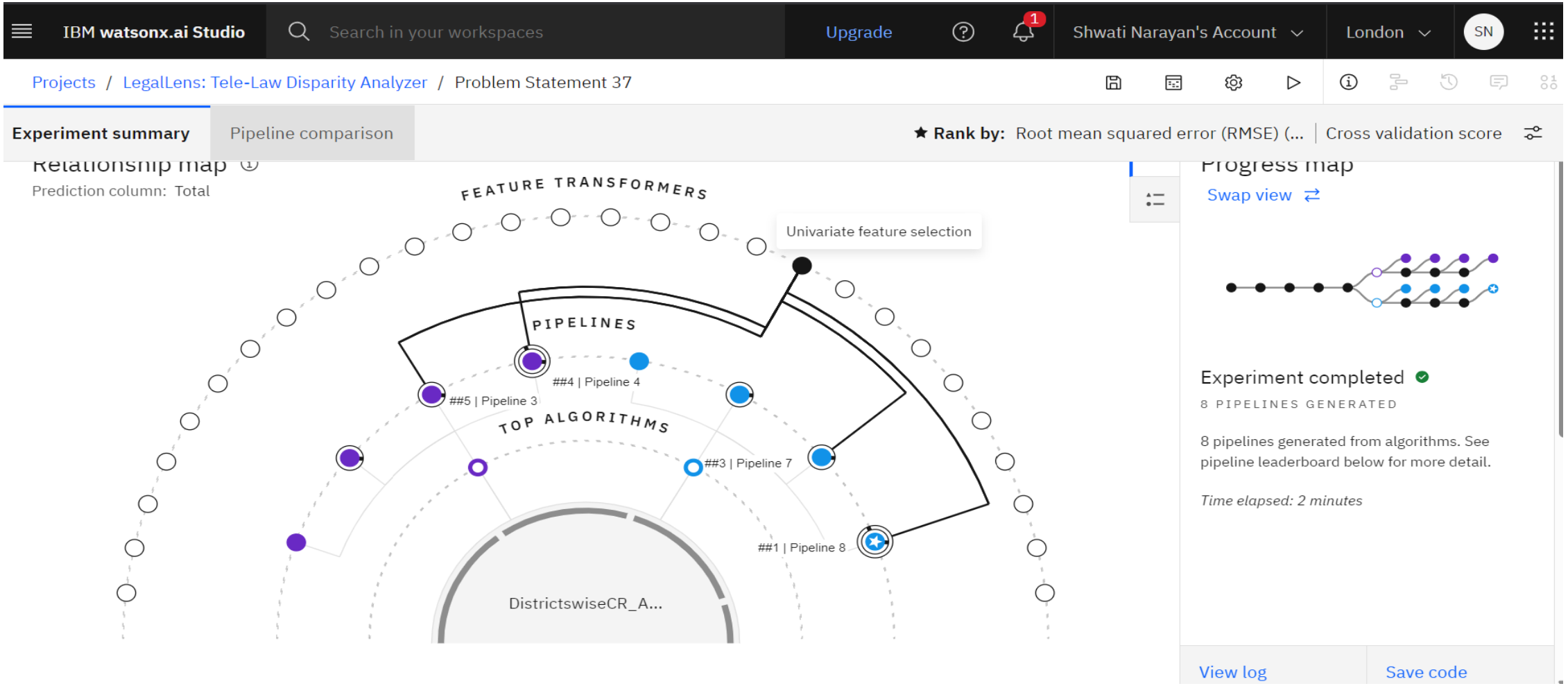
- **Algorithm Selection:**
  The project utilizes classification and regression algorithms such as Decision Trees and Random Forest, automatically selected via IBM AutoAI. These models are well-suited for structured tabular data and can handle complex patterns in categorical and numerical attributes like caste, gender, region, and CSC count. AutoAI ensures optimal model performance through automated hyperparameter tuning and model selection.

- **Data Input:**
  The model takes the following features as inputs:
  - Gender-wise case registration (Female, Male)
  - Caste categories (SC, ST, OBC, General)
  - No. of CSCs per district
  - District and State names (as categorical variables)

- **Training Process:**
  Historical Tele-Law data is cleaned and processed, then split into training and validation sets within IBM AutoAI. The platform applies automatic feature transformation, performs k-fold cross-validation, and tunes model parameters to maximize prediction accuracy.

- **Prediction Process:**
  The trained model predicts the total number of legal aid cases based on demographic and infrastructure variables. During prediction, real-time updates (like monthly data) can be fed to AutoAI for refreshed insights. Model outputs help identify underrepresented regions and marginalized communities in need of targeted outreach.

RESULT



edunet foundation

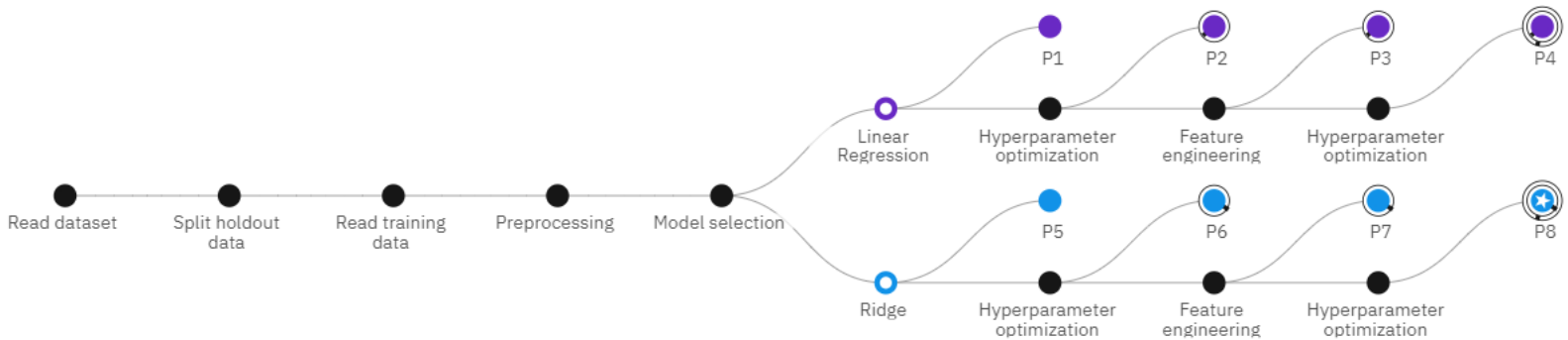
IBM watsonx.ai Studio | Search in your workspaces | Upgrade | 1 | Shwati Narayan's Account | London | SN

Projects / LegalLens: Tele-Law Disparity Analyzer / Problem Statement 37

Experiment summary | Pipeline comparison | ★ **Rank by:** Root mean squared error (RMSE) (... | Cross validation score

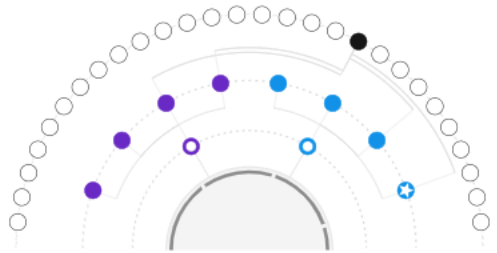
Progress map

Prediction column: Total



Relationship map

Swap view



Experiment completed

8 PIPELINES GENERATED

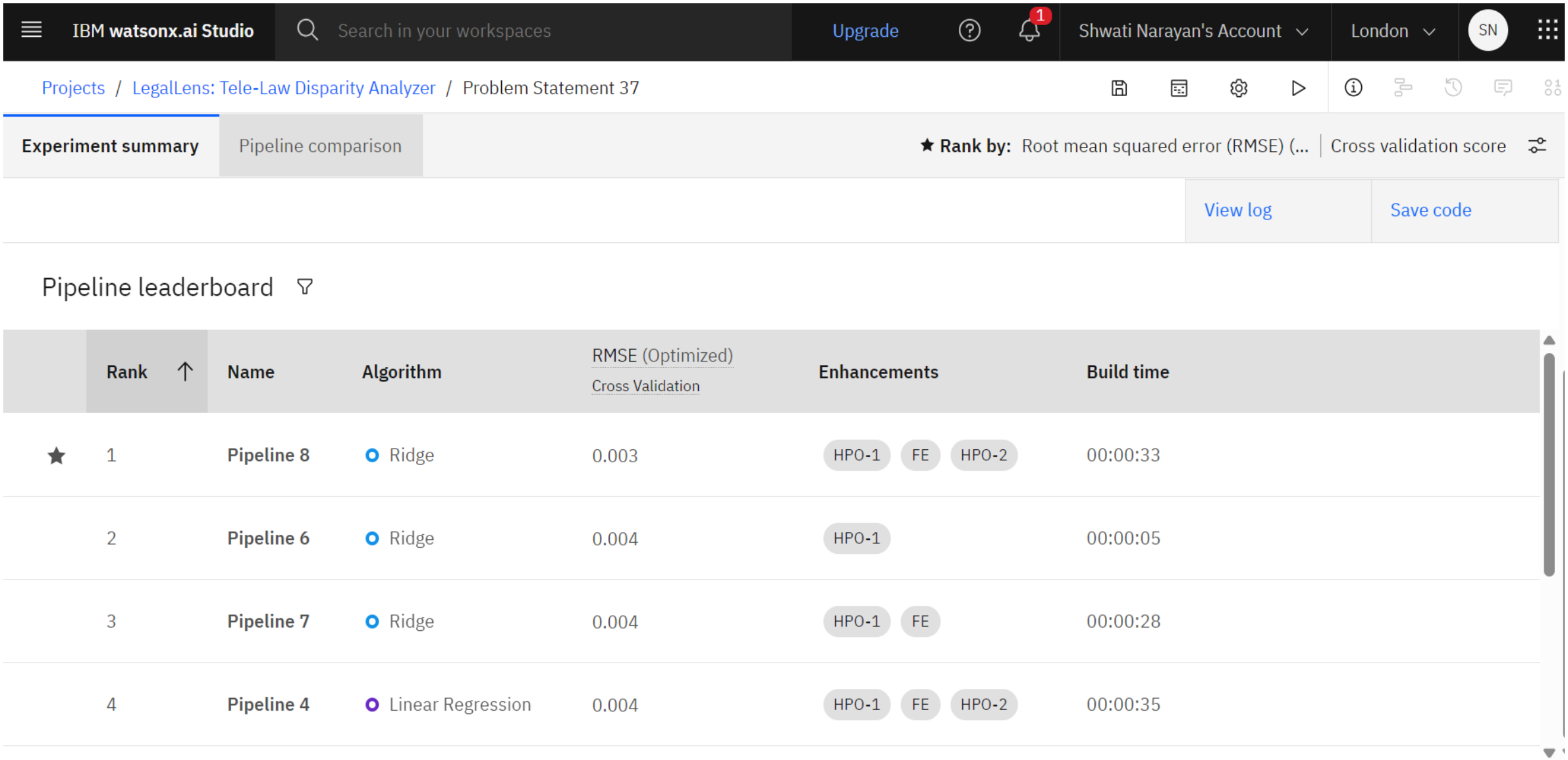
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 2 minutes

View log | Save code

Pipeline leaderboard

edunet foundation



IBM watsonx.ai Studio | Search in your workspaces | Upgrade | 1 | Shwati Narayan's Account | London | SN

Projects / LegalLens: Tele-Law Disparity Analyzer / Problem Statement 37

Experiment summary | Pipeline comparison

★ Rank by: Root mean squared error (RMSE) (... | Cross validation score

View log | Save code

Pipeline leaderboard

Rank	Name	Algorithm	RMSE (Optimized) Cross Validation	Enhancements	Build time
5	**Pipeline 3**	Linear Regression	0.004	HPO-1 FE	00:00:32
6	**Pipeline 2**	Linear Regression	0.006	HPO-1	00:00:08
7	**Pipeline 1**	Linear Regression	0.006	*None*	00:00:03
8	**Pipeline 5**	Ridge	0.006	*None*	00:00:02

IBM watsonx.ai Studio | Search in your workspaces | Upgrade | 2 | Shwati Narayan's Account | London | SN

Deployment spaces / newone / P8 - Ridge: Problem Statement 37 /

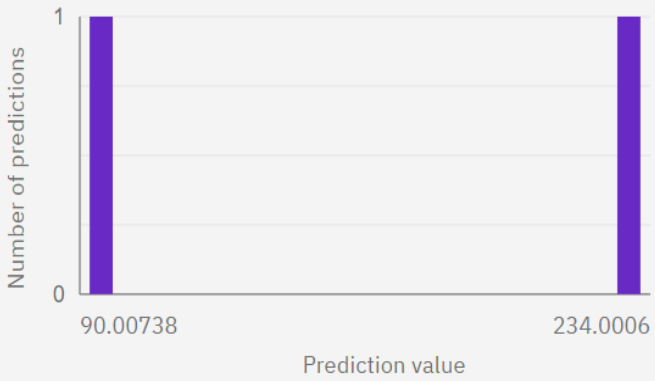
Prediction results

Close

Prediction type

Regression

Prediction distribution



Display format for prediction results

(•) Table view ( ) JSON view

Show input data

	Prediction	Category	States/UT's	Di
1	90.00738525390625	Case Registered	Arunachal Pradesh	Kr
2	234.0006866455078	Case Registered	Assam	M
3				
4				
5				
6				

Download JSON file

CONCLUSION

- The project successfully used IBM AutoAI to analyze Tele-Law registration data and identified clear demographic and regional disparities in legal aid access. Marginalized groups (SC, ST, OBC) and several districts showed low utilization despite having sufficient CSCs, highlighting a need for better outreach. The AutoAI model effectively predicted service usage patterns, enabling data-driven insights. Key challenges included ambiguous column names and data imbalance, which were addressed through preprocessing. Future improvements include integrating real-time data, normalizing by population, and using NLP for deeper analysis. Accurate predictions can help optimize legal resource distribution and promote inclusivity.

FUTURE SCOPE

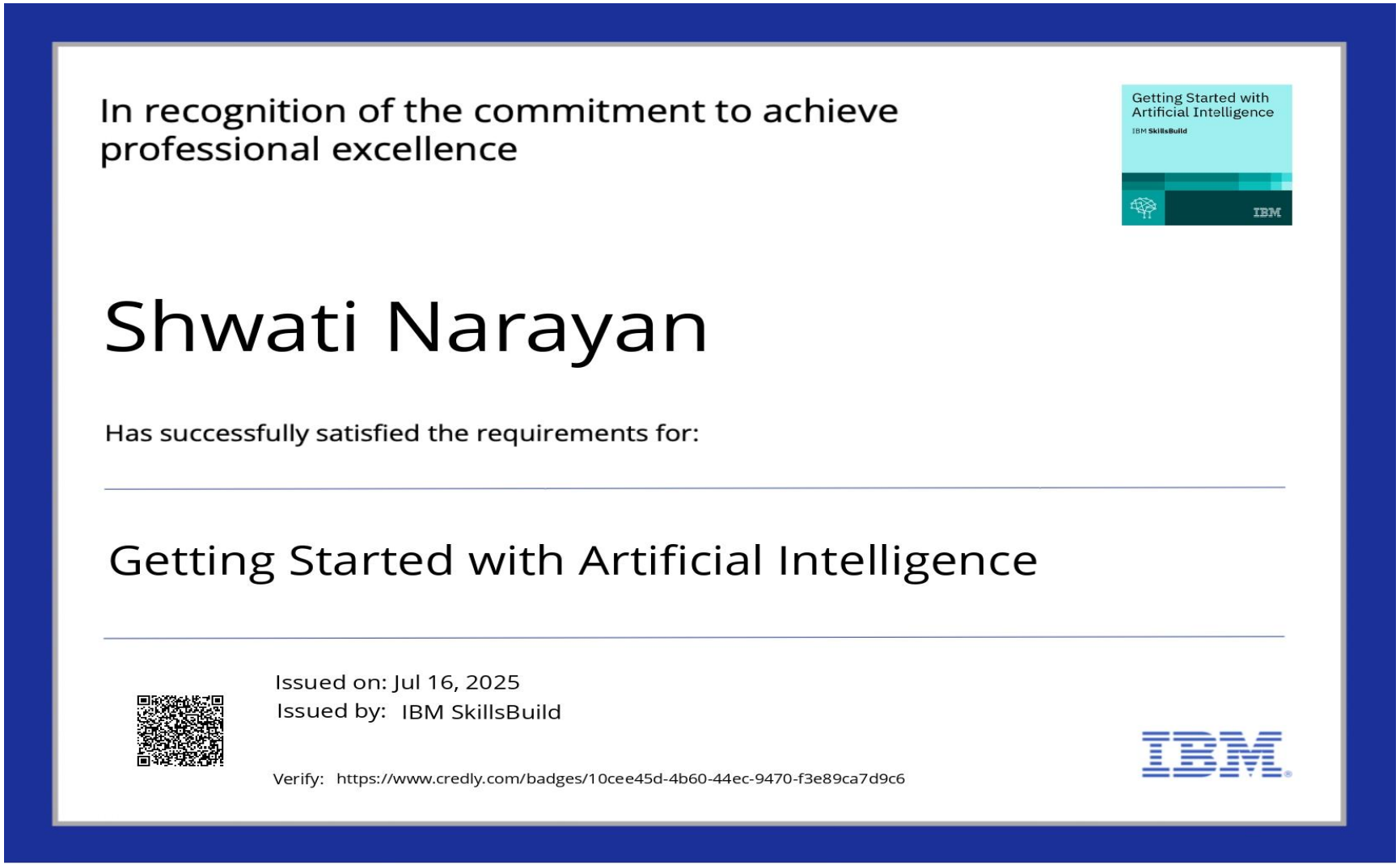
- **Integrate additional data sources**: Include demographic, socio-economic, and literacy rate data for deeper analysis.
- **Optimize model performance**: Use techniques like hyperparameter tuning, ensemble methods, or deep learning (e.g., LSTM, Transformers).
- **Geographic expansion**: Scale the system to cover multiple states or nationwide data for broader impact.
- **Adopt edge computing**: Enable real-time data processing directly at the Common Service Center (CSC) level.
- **Leverage advanced AI/ML**: Incorporate real-time recommendations and adaptive learning based on new data trends.
- **Enhance policy decision-making**: Provide actionable insights to improve Tele-Law outreach and inclusivity efforts.

REFERENCES

•[5] IBM Cloud Documentation. https://cloud.ibm.com/docs *(Reference for AutoAI, IBM Cloud Lite services, and deployment tools)*

•[6] Kaggle Dataset - **Bike Sharing Demand**: https://www.kaggle.com/c/bike-sharing-demand *(Historical bike rental data used in modeling)*

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

IBM

Shwati Narayan

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/10cee45d-4b60-44ec-9470-f3e89ca7d9c6

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

Shwati Narayan

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 21, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/81d9c7b8-2edf-426e-b615-b6a5ef6fa8d2

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Shwati Narayan

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 15 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet foundation