CAPSTONE PROJECT

MACHINE LEARNING PROJECT:- ANALYZING DEMOGRAPHIC AND REGIONAL DISPARITIES IN TELE LAW CASE REGISTRATIONS FOR INCLUSIVE LEGAL ACCESS

Presented By:
Sneha N
MVJ College of Engineering
CSE

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

Analyzing Demographic and Regional Disparities in Tele Law Case Registrations for Inclusive Legal Access

Despite the expansion of the Tele-Law initiative across states and districts, there is limited understanding of demographic utilization patterns and regional disparities in legal aid access. The challenge is to analyze Tele-Law case registration data to uncover gender-wise, caste-wise, and geographic disparities in service utilization across CSCs. Uneven representation among marginalized groups (SC, ST, OBC) and low outreach in certain districts raise concerns about equity and effectiveness. Moreover, the varying number of CSCs per region complicates direct comparisons. This problem demands a data-driven approach to evaluate inclusivity and optimize service delivery.

PROPOSED SOLUTION

The proposed system aims to address the challenge of ensuring equitable access to Tele-Law services across India. It focuses on identifying demographic (gender and caste) and regional disparities in legal aid utilization by analyzing Tele-Law case registration data. The solution involves leveraging data analytics and machine learning to uncover hidden patterns and suggest improvements. The system will consist of the following components:

- Data Collection:
  - Gather district-wise data on Tele-Law case registrations, including: Gender-wise (Male, Female), Caste-wise (SC, ST, OBC, General), Infrastructure-related (No. of CSCs).
  - Use additional sources if needed (e.g., district population, literacy rates) to enrich analysis.
- Data Preprocessing:
  - Clean and preprocess the dataset to handle any missing values or inconsistencies.
  - Derive new features such as:

    Gender Ratio = Female / Total

    SC/ST/OBC Ratios = SC or ST or OBC / Total

    Cases per CSC = Total / No. of CSCs
  - Normalize and scale the data for machine learning readiness.

- Machine Learning Algorithm:
  - Use clustering (e.g., K-Means) to group districts based on usage and inclusivity metrics.
  - Optionally apply regression to predict total case registrations using factors like CSC count and demographics.
  - Identify underutilized regions and patterns of disparity.
- Deployment:
  - Use IBM Watson Studio on IBM Cloud for model development and visualization.
  - Ensure accessibility and responsiveness for policy analysts and administrators.
- Evaluation:
  - Measure effectiveness using: Visual disparity detection, Cluster accuracy, Predictive accuracy metrics (if regression is used): MAE, RMSE
  - Iterate based on insights and feedback to improve model performance.

SYSTEM APPROACH

IBM Cloud: A scalable and secure platform for building and deploying AI/ML models.

- IBM Watson Studio: For building, training, and analyzing models using Python notebooks.
- IBM Cloud Object Storage: For storing the dataset securely in the cloud.
- IBM AutoAI: For automated model selection and hyperparameter tuning.
- IBM Cloud Pak for Data: For enterprise-level deployment, collaboration, and scalability.
- IBM Cloud Functions: For automating data processing or reporting tasks.

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  - The algorithm chosen is Linear Regression to predict the total number of Tele-Law case registrations in a district.
  - This algorithm was selected because the goal is to understand how features like number of CSCs, gender ratio, and caste representation influence total case counts — which is a continuous numeric output.
- **Data Input:**
  - The model uses the following features as input: No. of CSCs – infrastructure available, Female, Male – gender-wise case counts, SC, ST, OBC, General – caste-wise case counts.
  - Derived features (engineered during preprocessing):

    Gender_Ratio = Female / (Female + Male)

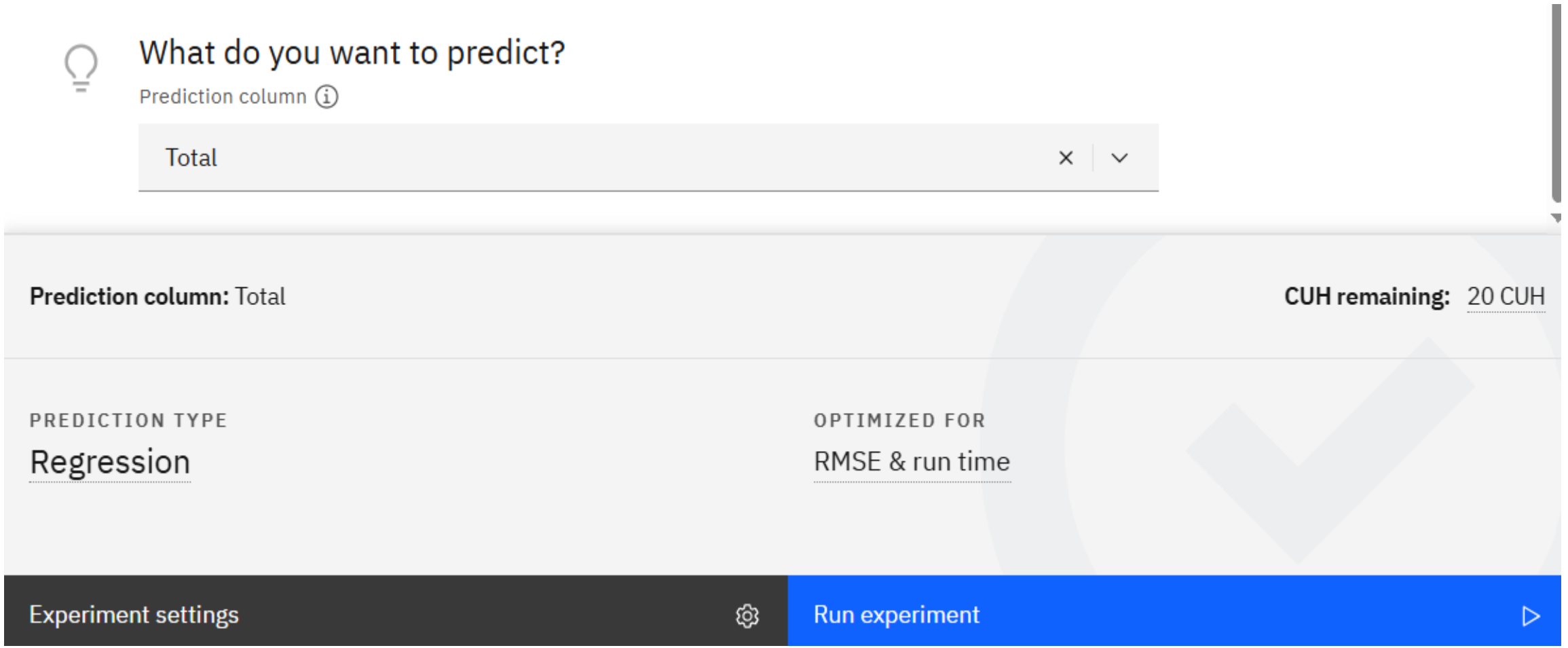
    Cases_per_CSC = Total / No. of CSCs

    Caste group ratios (e.g., SC_Ratio = SC / Total)

  These features were cleaned and structured inside IBM Watson Studio using an AutoAI pipeline.

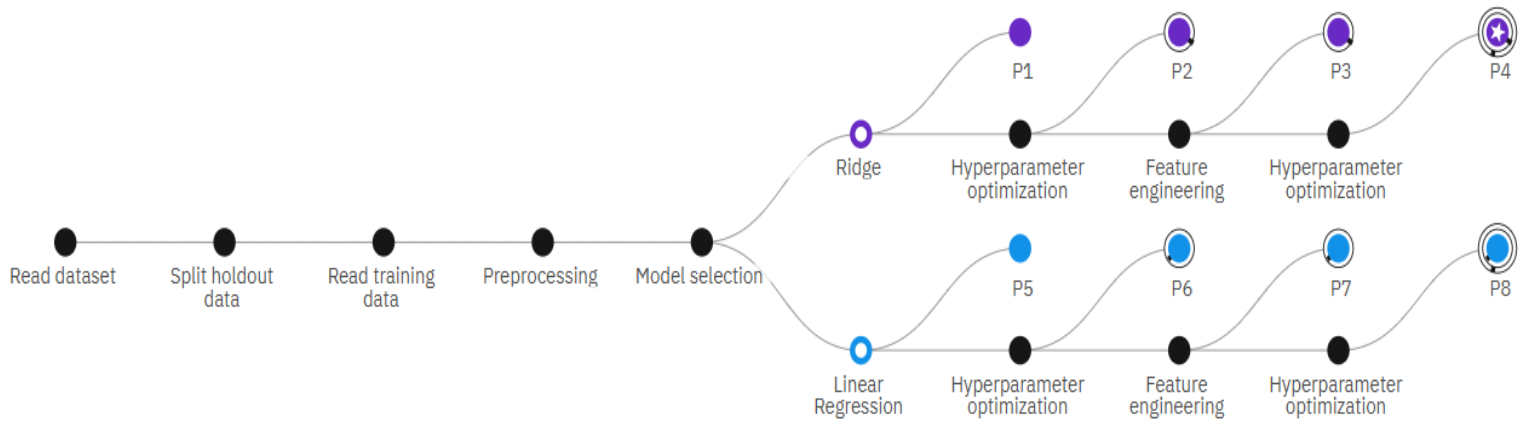
- **Training Process:**
  - The model was trained on historical case registration data using the regression algorithm inside IBM Watson Studio.
  - The data was split into training and testing sets to evaluate the model's generalization.
  - Model performance was assessed using regression metrics like RMSE (Root Mean Squared Error).
  - The training pipeline included automated feature selection and normalization.
- **Prediction Process:**
  - After training, the regression model was used to predict the expected number of Tele-Law cases for different districts.
  - A distribution graph showing how the predicted values are spread.These predictions can be compared with actual data to: Identify underperforming districts, Recommend outreach programs or policy interventions.

RESULT



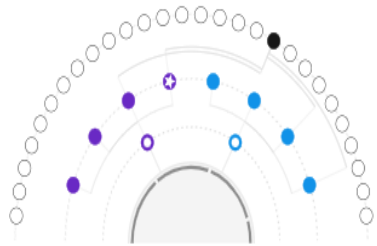
Progress map

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

	Rank	Name	Algorithm	RMSE (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 4**	Ridge	1.008e-8	HPO-1 FE HPO-2	00:00:42
	2	**Pipeline 3**	Ridge	1.008e-8	HPO-1 FE	00:00:37
	3	**Pipeline 2**	Ridge	0.004	HPO-1	00:00:15
	4	**Pipeline 8**	Linear Regression	0.004	HPO-1 FE HPO-2	00:00:33
	5	**Pipeline 7**	Linear Regression	0.004	HPO-1 FE	00:00:30
	6	**Pipeline 6**	Linear Regression	0.006	HPO-1	00:00:06
	7	**Pipeline 5**	Linear Regression	0.006	*None*	00:00:03
	8	**Pipeline 1**	Ridge	0.006	*None*	00:00:06

Deployment spaces / Tele_law_Deploy / P4 - Ridge: Tele_law_ML /

Tele_law_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

	Category (other)	States/UT's (other)	Districts (other)	No. of CSCs (integer)	Female (integer)	Male (integer)	General (integer)	OBC (integer)	SC (integer)	ST (integer)
1	Case Registered	Andaman and Nicot	Nicobar	5	615	852	557	315	546	49
2	Case Registered	Andhra Pradesh	Alluri Sitharama	430	6370	6828	3585	4660	3176	1777
3	Case Registered	Arunachal Pradesh	Lohit	32	87	84	133	10	19	9
4										
5										

3 rows, 11 columns

Predict

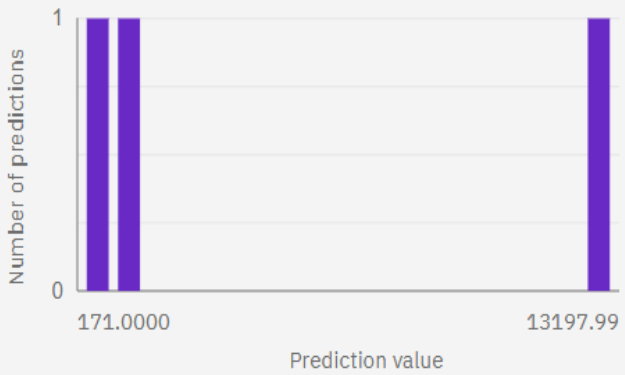
Prediction results

Close

Prediction type

Regression

Prediction distribution



Display format for prediction results

(•) Table view ( ) JSON view

Show input data

	Prediction
1	1467.0000000035545
2	13197.99999999832
3	171.00000000399797
4	
5	
6	
7	
8	
9	
10	
11	

CONCLUSION

The implemented system effectively analyzed Tele-Law case registration data using a regression model on IBM Cloud to identify disparities in legal aid access across districts. By predicting total case registrations based on factors like the number of CSCs, gender distribution, and caste representation, the model revealed significant usage gaps—particularly in districts with high infrastructure but low service uptake. Despite challenges such as data imbalance and the absence of population context, the solution provided actionable insights to support inclusive outreach planning. This project demonstrates how cloud-based analytics can guide equitable service delivery, much like accurate demand forecasting improves resource allocation in other domains.

FUTURE SCOPE

The system can be enhanced by integrating additional data sources such as district-wise population, literacy rates, internet penetration, and socio-economic indicators to better understand service gaps per capita. The prediction model can be optimized using advanced machine learning algorithms like Random Forest or Gradient Boosting for improved accuracy. Future versions could also include time-series analysis to monitor changes over months or years. The system can be scaled to cover more regions, including remote and tribal areas, and could incorporate real-time data pipelines using edge computing to track legal aid access as it happens. Additionally, deploying an interactive public dashboard will enable policymakers to take faster, data-driven actions.

REFERENCES

- Data set Link: https://www.data.gov.in/resource/district-wise-tele-law-case registration-and-advice-enabled-data-fy-2021-22-2024-25.
- IBM Documentation – Watson Studio and Cloud Pak for Data.
- IBM Cloud Docs – AutoAI, Cloud Functions, and Object Storage.

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Sneha N

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/ed48b3c3-1742-43d5-9a0e-847545b174d4

IBM

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

7/23/25, 6:48 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Sneha N

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU

edunet foundation