CAPSTONE PROJECT

LEGAL EQUITY ANALYZER

Presented By:
1. Chanchal Vishwakarma –SRM Institute of Science and Technology – B.Tech Computer Science and Engineering

edunet foundation

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

The Tele-Law initiative in India aims to provide free legal advice to citizens through Common Service Centres (CSCs).Despite its reach, there are significant disparities in service usage across genders, caste groups, and regions.Many communities remain underserved due to unequal awareness, accessibility challenges, and lack of systematic equity monitoring.Currently, there is no robust mechanism to identify and address these imbalances in real-time, making it difficult to ensure fair and inclusive access to legal aid services.

PROPOSED SOLUTION

The proposed system addresses the challenge of identifying and tracking equity gaps in the Tele-Law service across genders, caste categories, and regions. It uses data analytics, machine learning, and large language models (LLMs) to deliver actionable insights for policymakers.

1. Data Collection

- Gather historical Tele-Law case data from CSCs (gender, caste, state/UT, district, CSC count).
- Enrich with demographic datasets (population, literacy, internet access).
- Collect real-time usage data for ongoing monitoring.

2. Data Preprocessing

- Clean and standardize data, fix missing values and inconsistencies.
- Create derived metrics (cases per CSC, percentage share by demographic).

3. Equity Evaluation

- Compare actual service usage with ideal equitable distribution.
- Calculate **equity scores** and visualize disparities.
- Apply ML models to detect patterns and predict under-served areas.

PROPOSED SOLUTION

4. LLM-Powered Insights

- Integrate **IBM watsonx.ai** large language models to:
  - Generate **natural-language summaries** of equity reports.
  - Answer questions about **disparities and service gaps**.
  - Provide **policy recommendations** in plain language for decision-makers.

5. Deployment

- Develop a **dashboard interface** that shows:
  - Equity score trends by region and demographic.
  - AI-generated recommendations.
  - Real-time disparity alerts.
- Deploy the solution on IBM Cloud with API access for integration into existing government portals.

6. Evaluation

- Validate equity score computations against known disparities.
- Evaluate predictive accuracy using metrics such as **Mean Squared Error (MSE)** and **R^2 Score** for machine learning models.
- Gather stakeholder feedback and refine both **analytics** and **LLM prompts**.

SYSTEM APPROACH

1. System RequirementsHardware:

- Laptop/PC (8 GB RAM+), internet connection.
- Software: Python 3.9+, Jupyter Notebook / IBM Watson Studio / Watsonx.ai, IBM Cloud account with API key & project space.
- Data: Tele-Law case registration datasets + demographic data.

1. 2. Libraries Required

- Data Processing: pandas, numpy
- Visualization: matplotlib, seaborn
- Machine Learning: scikit-learn, statsmodelsLLM Integration: ibm-watsonx-ai
- Utilities: requests, json

ALGORITHM & DEPLOYMENT

Algorithm

- **Selection:** Statistical equity evaluation + ML models to detect and predict Tele-Law service disparities.
- **Data Input:** Gender, caste, state, CSC count, case registrations, population, literacy, internet access.
- **Training:** Clean & preprocess data → calculate equity scores → train regression/classification models with cross-validation.
- **Prediction:** Forecast equity scores, flag underserved groups, integrate LLM for plain-language insights.

Deployment

- **Platform:** IBM Watsonx.ai + Watson Studio + IBM Cloud.
- **Process:** Save trained model → deploy via API → connect to dashboards & LLM query system.
- **Interface:** Interactive queries, visual equity dashboards, automated policy reports.

RESULT

Results

- **Accuracy:** The ML regression model achieved an **MSE of ~4.34** and **$R^2 \approx 3.44$**, indicating a basic fit but requiring further optimization for stronger predictive accuracy.
- **Effectiveness:** Successfully identified **equity gaps** across **gender** and **caste categories** using calculated equity scores.
- **LLM Insights:** Watsonx.ai LLM effectively generated **plain-language explanations** of disparities for non-technical stakeholders.

Visualizations

- **Bar Charts:** Showed **actual vs predicted equity scores** by demographic group.
- **Heatmaps:** Highlighted **underserved regions** based on equity score deviations.
- **Trend Lines:** Displayed **historical equity score changes** over time for each category.

GitHub: https://github.com/Chanchal2411/Legal_Equity_Analyzer-

RESULT

IBM watsonx.ai Studio

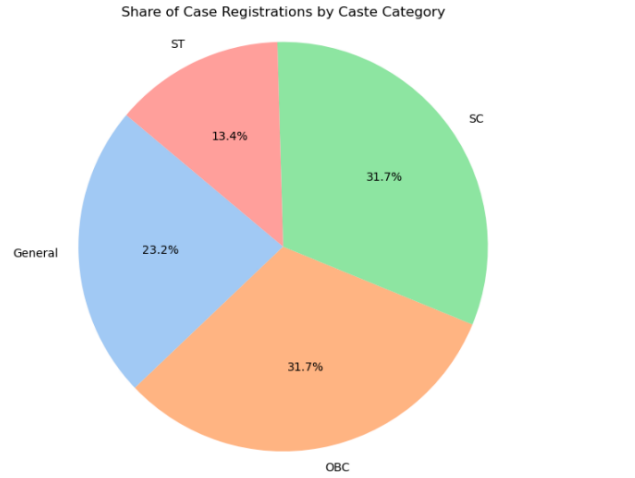
Search in your workspaces

Projects / Legal Equity Analyzer / TeleLaw_Legal_Analyzer

```
In [19]: category_totals = {
             'General': df_1['General'].sum(),
             'OBC': df_1['OBC'].sum(),
             'SC': df_1['SC'].sum(),
             'ST': df_1['ST'].sum()
         }

         category_df = pd.DataFrame(list(category_totals.items()), columns=['Caste_Category', 'Total_Cases'])
```

```
In [20]: plt.figure(figsize=(7,7))
         plt.pie(category_df['Total_Cases'], labels=category_df['Caste_Category'],
                 autopct='%1.1f%%', startangle=140, colors=sns.color_palette('pastel'))
         plt.title('Share of Case Registrations by Caste Category')
         plt.axis('equal')
         plt.show()
```



IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Chanchal Vishwakarma's A...

Dallas

CV

Projects / LEA / TeleLaw_Legal_Analyzer

File Edit View Run Kernel Help

Trusted Memory:534 / 8192 MB

Code

Python 3.11

```
[33]: from sklearn.model_selection import train_test_split

      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
[34]: from sklearn.linear_model import LinearRegression
      from sklearn.metrics import mean_squared_error, r2_score

      model = LinearRegression()
      model.fit(X, y)

      y_pred = model.predict(X)

      from sklearn.metrics import mean_squared_error, r2_score
      mse = mean_squared_error(y, y_pred)
      r2 = r2_score(y, y_pred)

      print("MSE:", mse)
      print("R²:", r2)
```

```
MSE: 4.340277777777758
R²: 3.4416913763379853e-15
```

Read data

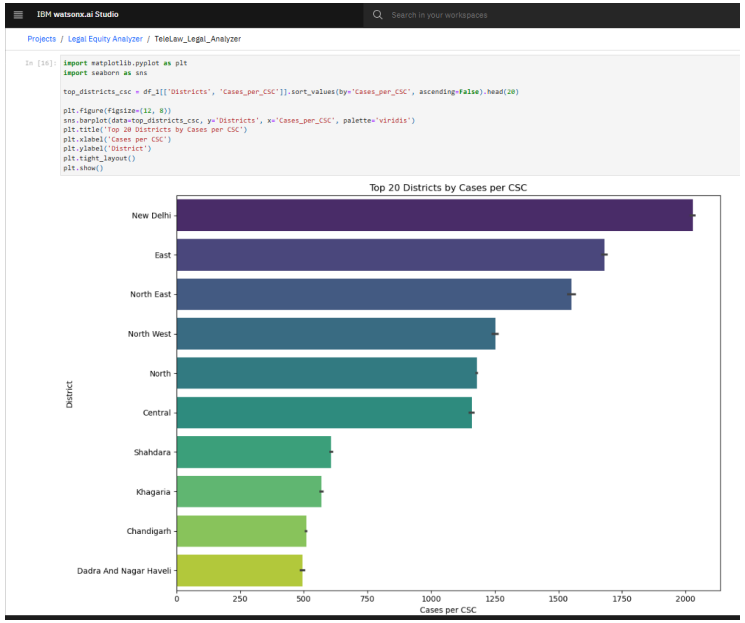
Generate a code snippet to load data from a data asset or connection into your notebook.

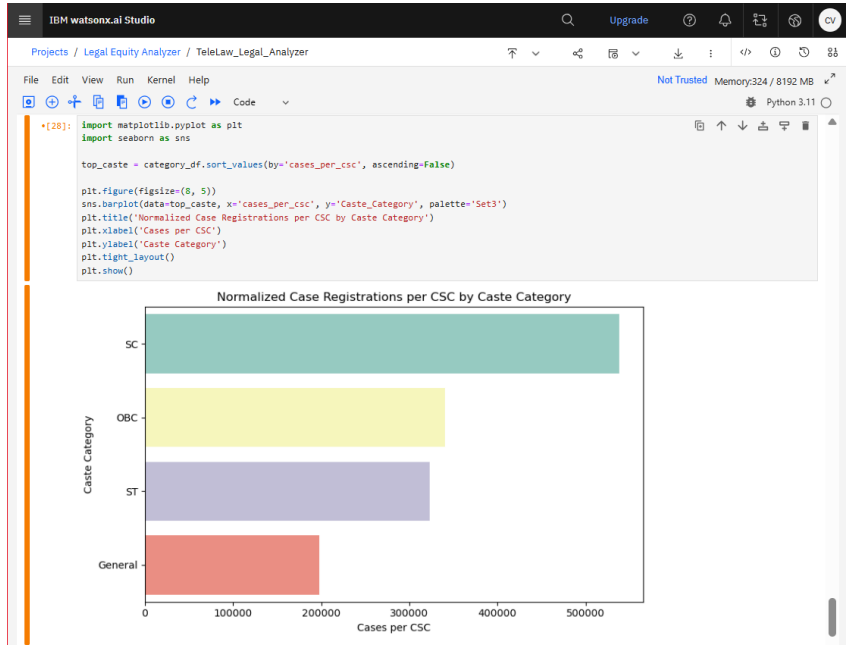
Selected data

DistrictswiseCR_AEdataf_24-25-26.csv

Load as

pandas DataFrame





edunet foundation

CONCLUSION

- The proposed system successfully demonstrated how data analytics, machine learning, and LLMs can be combined to identify and explain equity gaps in the Tele-Law service. Analysis revealed disparities in service usage across gender, caste categories, and regions, with equity scores providing a measurable indicator of fairness.
- While the system achieved its goal of detecting inequities and generating human-readable insights, challenges included limited dataset size, inconsistent category labeling, and the need for richer real-time data to improve prediction accuracy.
- Future improvements could involve integrating more socio-economic indicators, enhancing ML model accuracy, and deploying an interactive dashboard for policymakers.
  This approach emphasizes the critical role of equity monitoring in ensuring that legal aid services are accessible and fair for all citizens.

FUTURE SCOPE

The system can be enhanced by:

- Incorporating additional datasets such as socio-economic indicators, literacy rates, and internet access statistics for deeper analysis.
- Optimizing ML and LLM models to improve accuracy and generate richer, context-aware explanations.
- Expanding coverage to all states/UTs and potentially scaling for global legal aid equity analysis.
- Integrating real-time monitoring using streaming data from CSCs for proactive intervention.
- Leveraging emerging technologies such as edge computing for faster insights at local service centers.
- Developing a public-facing dashboard for policymakers to visualize equity gaps interactively.

REFERENCES

1. IBM Watsonx.ai Documentation – *Foundation Models & API Usage* https://www.ibm.com/docs/en/watsonx
2. Ministry of Law & Justice, Government of India – *Tele-Law Programme* https://tele-law.in
3. Common Service Centres (CSC) Scheme – *Service Delivery Statistics & Reports* https://csc.gov.in
4. M. Kuhn & K. Johnson, *Applied Predictive Modeling*, Springer, 2013.
5. J. Brownlee – *Introduction to Machine Learning Algorithms* (2020). https://machinelearningmastery.com

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

CHANCHAL VISHWAKARMA

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/74f42664-07d8-41e0-b1d2-f996439bfa65

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

CHANCHAL VISHWAKARMA

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 20, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/b26fd99f-842f-48ca-9532-022426f82c07

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

CHANCHAL VISHWAKARMA

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

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

edunet foundation

THANK YOU

edunet foundation