



RV College of Engineering

Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

**9 INDUSTRY, INNOVATION
AND INFRASTRUCTURE**



Go, change the world

V SEMESTER

Experiential Learning - ODD 2025-26

Phase-I Review

on

Theme: SDG 9 (Computer Science)
**SENTIMENT ANALYSIS OF PUBLIC
FEEDBACK ON INFRASTRUCTURE**



Team Members

PIYUSH KHERIA	AADITYA RAJ	SHRIYANSH AGRAWAL	OOJAM CHAUDHARY	SAMKIT SAMSUKHA
1RV23CS166	1RV23CY001	1RV23CY064	1RV23IS079	1RV23IS105
piyushkheria.cs23@rvce.edu.in	aadityaraj.cy23@rvce.edu.in	shriyansha.cy23@rvce.edu.in	oojamchaudhary.is23@rvce.edu.in	samkitsamsukha.is23@rvce.edu.in

MENTOR

Dr. Sindhu D V
Assistant Professor
Department of Computer Science and Engineering
RV College of Engineering
Bengaluru-59



RV College of Engineering

Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

Go, change the world

Contents

1. Introduction
2. Existing system
3. Identification of problem
4. Literature Review
5. Problem definition
6. Objectives
7. Methodology
8. Proposed tools and techniques.
9. Expected outcome.
10. Conclusion
11. References



Introduction

- Public infrastructure projects generate vast feedback across portals, social media, and government platforms.
- Sentiment analysis converts this unstructured data into meaningful, measurable insights.
- It helps identify public opinion, recurring issues, and satisfaction levels.
- These insights support planners and policymakers in making informed, citizen-centric decisions.

**50 million citizen
feedback posts every
month**

**72% of this
data is
unstructured**

60% of citizens express infrastructure-related frustrations online, yet less than 20% of this feedback is currently analysed in a structured manner



Existing System In India

- PM GatiShakti Public Platform**

A national masterplan portal used for infrastructure planning, offering access to government-approved datasets and analytics. It integrates feedback mechanisms, allowing users (including private sector and citizens) to analyse infrastructure sentiment and design optimal projects

- MyGov Platform**

The Government of India's citizen engagement portal collects feedback and opinions about schemes and infrastructure, using data science tools for public sentiment analysis across millions of users



Existing System

- **Twitter-based Sentiment Analysis**

Research frameworks in Indian academia leverage location-based social network data (primarily Twitter) for sentiment tracking on metro, roads, and smart city infrastructure, identifying commuter pain points and preferences.

- **Sentiment Classification on Government Schemes**

Academic implementations apply multidimensional sentiment analysis using systems like PySpark, deploying models on feedback about infrastructure schemes to classify user emotion across several categories



Existing System Global Solutions

- **IBM Watson Natural Language Understanding**

Provides APIs for sentiment analysis across languages, allowing government and private entities to monitor citizen feedback on infrastructure projects globally with standardized NLP frameworks

- **Brandwatch**

Used for multidimensional sentiment analysis on global infrastructure feedback, capable of tracking trends and opinions at scale, including entity-level and regional breakdowns



Identification of problem

- **Small and Imbalanced Datasets:** Datasets are limited in size and skewed toward positive reviews, causing models to overfit and fail to represent broader public opinion.
- **Lack of Local Context and True User Perspective:** Reliance on global social media distorts real infrastructure-related feedback, as local, ground-level user voices remain underrepresented.
- **Limited Coverage of Negative Sentiment and Specific Aspects:** Positive and neutral feedback dominates most datasets, resulting in poor visibility of critical issues and actionable negative insights.



A similar problem statement was also featured in the latest Smart India Hackathon, highlighting its rising relevance and trend.



Literature Review

Sentiment Analysis for Citizen Feedback in Smart Cities with XLNet-BiLSTM (Delhi Metro case)

It uses **XLNet** for contextual embeddings and **BiLSTM** for sequence understanding, trained first on the **IMDb dataset (93.1% accuracy)** and later tested on a 50-review custom dataset, achieving **100% accuracy —likely due to overfitting given small, imbalanced data.** Tools used include XLNet, BiLSTM, tokenization, preprocessing, confusion matrices, and accuracy/loss plots. The study highlights the potential of sentiment analysis for smart city decision-making and calls for larger, diverse datasets.

Sentiment Analysis of Public Services for Smart Society: Literature Review and Future Directions

It highlights how SA—through NLP, machine learning, and text mining—supports citizen-centric governance, policy formulation, and real-time social insight. **The study analyzes research trends since 2010**, revealing fragmented, cross-disciplinary literature. Using conceptual, thematic, and co-citation analyses, it maps the field's intellectual structure and identifies emerging themes. The paper concludes that **SA is essential for smart governance but requires context-specific models, stronger integration with civic technology, and expanded future research.**



Literature Review

Decoding Public Sentiment Topics in Google Map Reviews for BRI Urban Infrastructure Projects

This study analyzes **144,210 Google Maps reviews from 352 Belt and Road (BRI)** urban infrastructure projects to understand public perception using VADER sentiment analysis and Mg-LDA topic modeling. **Unlike media or Twitter data, Google reviews offer direct, location-specific user experiences.** Four major discussion themes emerge: professional function, service industry, benefits, and development.

Sentiment Analysis on Government Public Policies: A Systematic Literature Review

It analyzes 27 high-quality studies and finds **SVM, BERT, Naive Bayes, and Random Forest as the most effective methods**, while VADER remains popular for quick lexicon-based analysis. Twitter is the dominant data source, followed by YouTube and Weibo. Evaluation commonly uses confusion matrix, precision, recall, and F1-score. Research trends have shifted from health-related policies to infrastructure, environment, and technology



Literature Review

Unveiling the Impact of Indian Government Policies Using Aspect-Based Sentiment Analysis

This study analyzes public sentiment toward Indian government policies using **Aspect-Based Sentiment Analysis (ABSA), Multi-Criteria Decision-Making (MCDM), and a hybrid deep-learning model combining CNN, Bi-LSTM, and Transformers**. Using 310,786 policy-related tweets, it evaluates sentiments across aspects like economic reforms, welfare, healthcare, education, and environment. The study links policy sentiment with electoral mood, offering insights for policymakers ahead of India's 2024 elections.

[Deena Nath. (2024). Unveiling the Impact of Indian Government Policies using Aspect Based Sentiment Analysis with Multi-Criteria Decision-Making and Hybrid Deep Learning. International Journal of Intelligent Systems and Applications in Engineering, 12(22s), 72–82. Retrieved from <https://www.ijisae.org/index.php/IJISAE/article/view/6395>]

Survey on Sentiment Analysis: Evolution of Research and Application in Smart Societies

This survey analyzes **9,714 sentiment-analysis papers (2002–2022)** to map how research methods and topics have evolved. Using keyword co-occurrence, Louvain community detection, and tools, the study identifies six major research communities, covering **social media analytics, machine learning, deep learning, opinion mining, Arabic SA, and cross-domain methods**. Findings show a shift from early machine-learning and lexicon methods toward **deep learning, aspect-based analysis, multilingual SA, and cross-domain adaptation**.

[Cui J, Wang Z, Ho SB, Cambria E. Survey on sentiment analysis: evolution of research methods and topics. Artif Intell Rev. 2023 Jan 6:1-42. doi: 10.1007/s10462-022-10386-z. Epub ahead of print. PMID: 36628328; PMCID: PMC9816550.]



Problem Definition

Current sentiment-analysis systems for public infrastructure planning face critical shortcomings that limit their reliability and real-world applicability. They rely on **small, imbalanced, and fragmented datasets** that fail to reflect actual citizen experiences. The heavy dependence on global social media further ignores local, infrastructure-specific feedback, leading to **under-representation of genuine on-ground issues**. Additionally, multilingual diversity, sarcasm, code-mixing, and contextual nuances commonly found in public discourse are poorly handled, resulting in inaccurate sentiment classification. These gaps collectively lead to weak insights and hinder evidence-based decision-making for planners and policymakers. Addressing these challenges requires building a **comprehensive, multilingual, and context-aware citizen sentiment intelligence platform for public infrastructure**.



Objectives

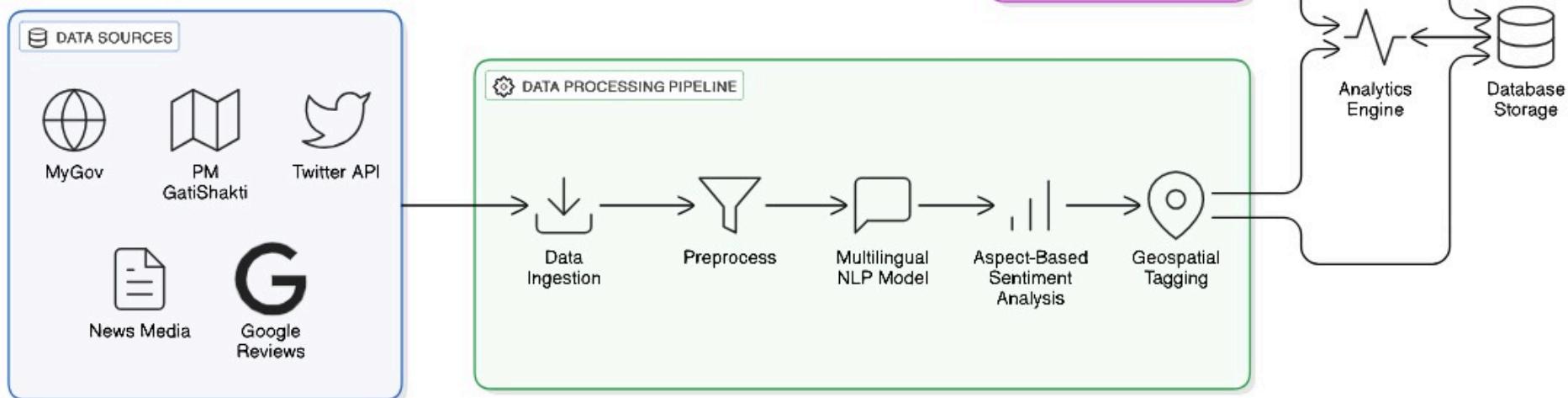
- Build a **large, diverse, and balanced citizen feedback dataset** on public infrastructure by combining multiple channels such as portals, social media, reviews, and surveys.
- Ensure **fine-grained, aspect-based, and location-aware sentiment analysis** (e.g., road quality, safety, cleanliness, public transport) instead of relying only on overall sentiment.
- Improve **multilingual accuracy for Indian languages and dialects** using modern cross-lingual models and high-quality annotation.
- Maintain **reliability** using explainable and robust models capable of handling sarcasm, code-mixing, and noisy text.
- Build an **actionable decision-support dashboard and alert system** that highlights issues, locations, and sentiment trends over time.



Methodology

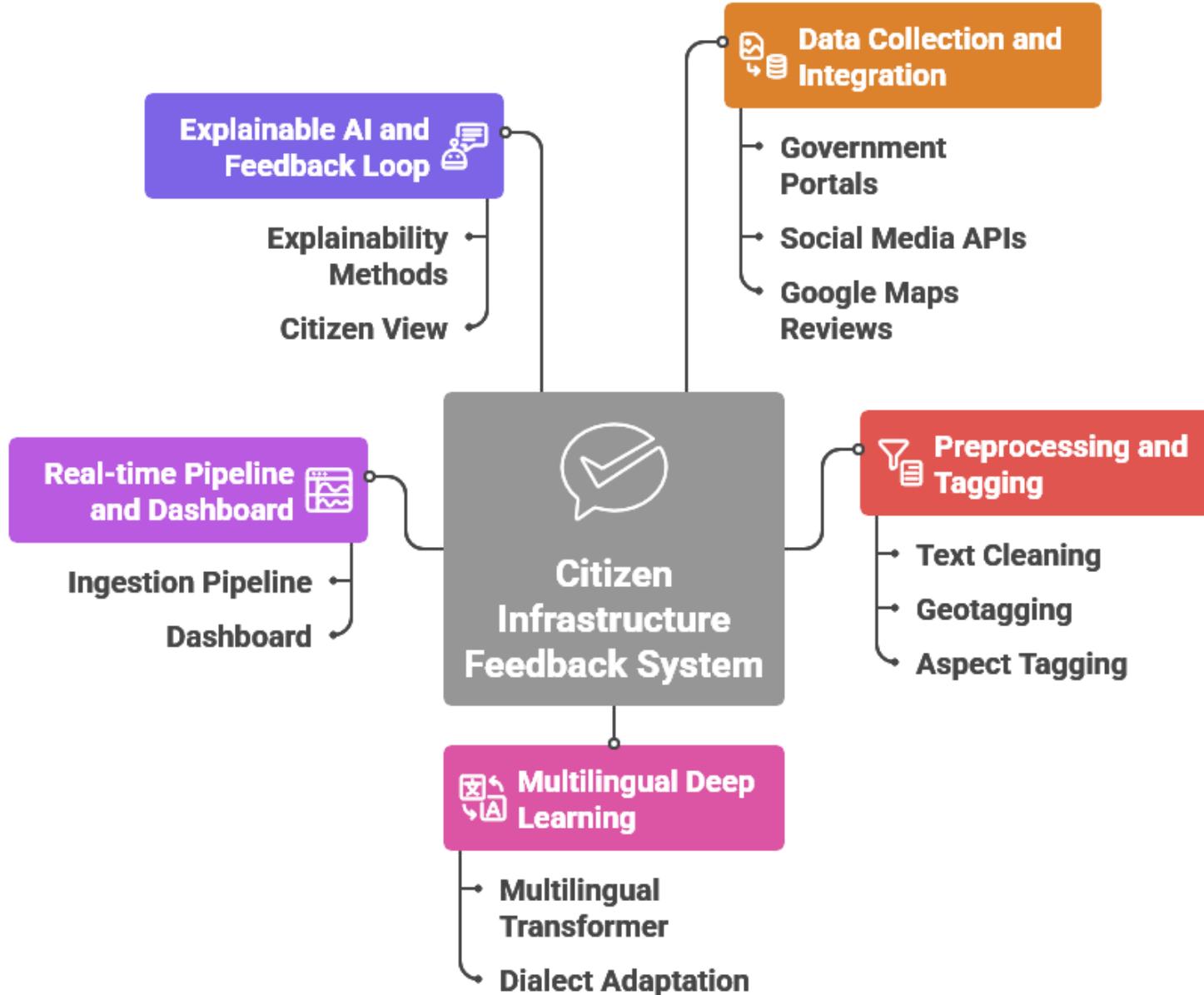
Go, change the world

The system collects citizen feedback from multiple sources such as MyGov, PM GatiShakti, Twitter, news media, and Google Reviews, which are fed into a unified data-processing pipeline. After ingestion, the data is cleaned, normalized, and preprocessed to handle multilingual text and Indian code-mixed language. A custom multilingual NLP model performs aspect-based sentiment analysis to extract sentiments tied to specific infrastructure themes like transport, cleanliness, safety, or public services. Each processed data point is then enriched with geospatial tags, enabling location-aware insights. The analytics engine stores and organizes this enriched data for interactive dashboards used by policymakers, analysts, admins, and even the general public, allowing them to explore trends, compare regions, and identify hidden issues through temporal and spatial visualizations.



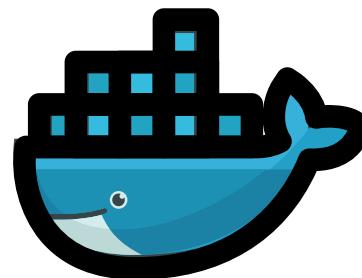
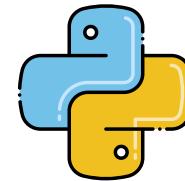


Methodology





Proposed tools and techniques



Hugging Face

MuRIL

ai4bharat

IndicLID

IndicBERT

Captum



Expected Outcomes

Objective	Expected Outcome
1. Build a large, diverse, balanced citizen feedback dataset.	A unified dataset from multiple channels covering major cities with balanced sentiment classes and strong representation of ward-level complaints.
2. Enable fine-grained, aspect-based, location-aware sentiment analysis.	Ward-wise sentiment heatmaps and automatic detection of specific issues (e.g., potholes, garbage hotspots, bus delays).
3. Improve multilingual accuracy for Indian languages/dialects.	Higher F1-scores across major Indian languages and correct interpretation of dialectal phrases and code-mixed text.
4. Use explainable, robust models for sarcasm, code-mixing, and noisy text.	Accurate detection of sarcastic/ironic feedback with clear model explanations showing why each sentiment was assigned.
5. Build an actionable dashboard and alert system.	Real-time alerts for sentiment spikes in specific localities and trend analytics that show issue severity and improvement over time.



Conclusion

The project will continue to evolve into a unified, multilingual platform that transforms raw citizen feedback into actionable insights for improving public infrastructure across India. It will build a large, diverse dataset from multiple channels, develop fine-grained and location-aware sentiment models, and strengthen multilingual performance through advanced cross-lingual techniques. The system will also incorporate explainable and robust NLP methods capable of handling code-mixing, sarcasm, and noisy real-world text. As development progresses, the project will deliver an interactive decision-support dashboard and alert system that highlights emerging issues, tracks sentiment trends over time, and empowers policymakers, analysts, and administrators with data-driven intelligence. Ultimately, the project will contribute toward more transparent, responsive, and citizen-centric infrastructure planning.



References

1. CEUR Workshop Paper on Infrastructure Feedback
2. ScienceDirect – Public Infrastructure Sentiment Dataset Study
3. TechUK – Location-Aware Public Service Innovation
4. WorldComp Proceedings – Decision-Support Dashboard Paper
5. Mihup – Techniques to Improve Multilingual Sentiment Accuracy
6. Nature – Modern Sentiment Analysis Advances
7. ScienceDirect – Robust Multilingual SA Models (Sarcasm & Noisy Text)
8. InteresJournals – SA Techniques, Limitations & Case Studies
9. DataDrivenInvestor – Sentiment Analysis for Smart Cities
10. PMC – Research Article on Sentiment Analysis Methods
11. RSIS International – Sentiment Analysis Case Study (India)
12. ScienceDirect – SA Applications in Public Policy & Services



RV College of Engineering®

Mysore Road, RV Vidyaniketan Post,
Bengaluru - 560059, Karnataka, India

Go, change the world

Thank You