Civic Sentiment Analysis Platform – MVP

Problem Statement

Citizens share civic complaints (traffic, water, sanitation, waste) on social media and portals, but authorities lack a real-time automated system to analyze them. Manual methods are slow and inconsistent. We need an AI-based platform to **analyze sentiments, group issues, and provide explainable insights** via dashboards.

Objectives (MVP)

1. Collect & preprocess civic feedback data (cleaning, tokenization, lemmatization).
2. Implement **sentiment analysis** with pretrained transformers (BERT/DistilBERT) + **Explainable AI (XAI)** for interpretability (SHAP/LIME).
3. Apply **topic modeling** (LDA/BERTopic) to group complaints into domains (traffic, water, sanitation).
4. Develop an **interactive dashboard** (Flask/Django + Plotly/Chart.js) to visualize results and explanations.

Methodology (Pipeline)

- **Data Collection** → social media APIs / civic portals / sample datasets.
- **Preprocessing** → text cleaning, stopword removal, lemmatization, translation.
- **Sentiment Analysis** → BERT/DistilBERT (Positive, Neutral, Negative).
- **XAI** → SHAP/LIME/Attention → explain predictions.
- **Topic Modeling** → LDA & BERTopic for issue categorization.
- **Dashboard** → visualize sentiment trends, topic distribution, and explanations.

Tools & Technologies

- **Python** → core programming language.
- **NLTK, SpaCy, Hugging Face** → preprocessing & transformers.
- **PyTorch/TensorFlow** → ML model support.
- **SHAP, LIME** → Explainable AI.
- **LDA, BERTopic** → topic modeling.
- **SQLite/PostgreSQL** → database.
- **Flask/Django + Plotly/Dash/Chart.js** → backend & visualization.
- **Streamlit/Heroku** (optional) → simple deployment.

MVP vs Future Scope

MVP (This Semester)	Future Scope (Next Phases)
Data from sample civic datasets & small APIs	Large-scale multi-city, real-time streaming
Basic preprocessing (cleaning, stopwords, lemmatization)	Advanced multilingual + sarcasm detection
Sentiment (3 classes) + XAI (SHAP/LIME)	Multi-emotion analysis + advanced interpretability
LDA + BERTopic for topics	Hierarchical/dynamic topic modeling
Basic dashboard with charts	Advanced dashboards with maps, forecasting
SQLite/Postgres	Scalable DB (MongoDB, Hadoop)
Localhost/Heroku demo	Cloud-native deployment + chatbot/voice support