Mini-Project – 2B Web based on ML (ITM 601)

Live Insights on General Mood and Attitude

T. E. Information Technology

By

Tanmay Bhatkar 09
Mazin Bangi 08
Sahil Bangera 07
Shannen Anthony 04

Under Guidance of
Dr. Minal Lopes
Assistant Professor

Department of Information Technology
St. Francis Institute of Technology
(Engineering College)
(An Autonomous Institute, Affiliated to University of Mumbai)
AICTE Approved | NAAC A+ grade: 03 programs are NBA Accredited: ISO 9001:2015 Certified

2024-2025

CERTIFICATE

This is to certify that the project entitled **"Live Insights on General Mood and Attitude"** is a bonafide work of **"Tanmay Bhatkar,Mazin Bangi, Sahil Bangera, Shannen Anthony" Roll Nos. 09, 08, 07, 04** submitted to the University of Mumbai towards completion of mini project work for the subject of **Mini Project – 2B Web Based on ML (ITM 601).**

Dr. Minal Lopes
Mini Project Guide

Dr. Prachi Raut
HOD-INFT

Examiners

1.---

2.---

Date:

DECLARATION

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

--------------------------------------	--------------------------------------
Tanmay Bhatkar	Mazin Bangi
--------------------------------------	--------------------------------------
Sahil Bangera	Shannen Anthony

ABSTRACT

Effective policymaking requires real-time insights into public sentiment, especially in a diverse nation like India. Traditional methods such as surveys and polls are often outdated, biased, or unreliable. To address this, we propose LIGMA (Live Insights on General Mood & Attitudes)—a sentiment analysis system that dynamically interprets public opinion using real-time data from X (formerly Twitter) and news sources. LIGMA employs a dual BERT-based machine learning approach: a multi-label sentiment classifier fine-tuned on GoEmotions and SemEval for nuanced emotion detection on news articles, and a fine-tuned BERT model on the Twitter Emotions Corpus to accurately categorize public opinions expressed in tweets. For data extraction LIGMA scrapes top news results from Google.com, applying cosine similarity and TF-IDF-based relevance scoring to filter high-quality articles based on topic, which are then stored in MongoDB. Similarly, tweets on the same topics are scraped via Selenium, preprocessed using NLP techniques including tokenization and summarization, and analyzed for sentiment. The system operates on a real-time streaming architecture with continuous data ingestion and processing. Results are visualized through interactive dashboards featuring trend analysis, word clouds, and fine-grained sentiment insights. By mitigating bias and enhancing sentiment accuracy, LIGMA provides policymakers with transparent, data-driven public sentiment analysis, enabling more informed and responsive governance.

Index

List of Abbreviations

Sr. No.	Abbreviation	Full Form
1	LIGMA	Live Insights on General Mood and Attitude
2	TF-IDF	Term Frequency-Inverse Document Frequency
3	NLP	Natural Language Processing
4	NSSO	National Sample Survey Office
5	LLM	Large Language Model
6	AI	Artificial Intelligence
7	MCDM	Multi-Criteria Decision Making
8	LSTM	Long Short-Term Memory
9	CNN	Convolutional Neural Network
10	F-WOA	Fuzzy Whale Optimization Algorithm
11	PCA	Principal Component Analysis
12	HDNN	Hierarchical Deep Neural Network
13	SVM	Support Vector Machine
14	RMSE	Root Mean Square Error
15	RoBERTa	Robustly Optimized BERT Pretraining Approach
16	VADER	Valence Aware Dictionary and sEntiment Reasoner
17	CSV	Comma-Separated Values
18	JSON	JavaScript Object Notation
19	URL	Uniform Resource Locator
20	GTX	Giga Texel Shader eXtreme
21	GPU	Graphics Processing Unit
22	GDPR	General Data Protection Regulation
23	DPDP	Digital Personal Data Protection (Act, India)

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

Public sentiment has long been a cornerstone of governance, yet history is replete with policies that failed due to a disconnect from societal attitudes. The 1990s British poll tax, which provoked mass unrest and eventual repeal, exemplifies the consequences of ignoring grassroots dissent. Similarly, India's 2020 Farm Bills faced nationwide protests over perceived exclusion of farmer input, while vaccine hesitancy during COVID-19 revealed how delayed insights into public trust can hinder crisis response. These cases underscore a persistent challenge: traditional mechanisms for gauging public opinion—such as surveys, polls, and focus groups—are often inadequate for capturing real-time, nuanced sentiment.

Surveys, while structured, suffer from inherent delays, limited sample sizes, and participation biases (e.g., underrepresenting rural or digitally excluded populations). For instance, India's National Sample Survey Office (NSSO) releases data months after collection, rendering it obsolete for rapid policy adaptation. Polls, too, are vulnerable to framing effects and oversimplification, reducing complex emotions to binary choices. Even modern alternatives like town halls or letters to officials lack scalability and risk amplifying vocal minorities over silent majorities.

The digital age offers a paradigm shift. Social media platforms like X (Twitter) and news outlets generate vast, organic data reflecting real-time public discourse. However, existing sentiment analysis tools struggle to parse this deluge effectively. Many rely on simplistic polarity models (positive/negative), ignoring context-specific emotions (e.g., fear, trust, or ambivalence) vital for policymaking. Others conflate media narratives with genuine public opinion or fail to filter noise from relevant discourse.

These gaps highlight the need for systems that dynamically synthesize multi-source data, decode layered emotions, and deliver timely insights—without the lag or bias of legacy methods. Such tools could bridge the chasm between governance and societal needs, transforming reactive policymaking into proactive, inclusive stewardship.

1.2 Need and Scope of the project

Government policies directly shape citizens' lives, from healthcare access to economic stability. Informed decision-making hinges on understanding public sentiment, as misaligned policies—like the 1990s British poll tax or India's contentious Farm Bills—risk social unrest, economic losses, and eroded trust. Conversely, policies grounded in public needs, such as participatory budgeting initiatives in Kerala, India, demonstrate how inclusive governance fosters accountability and civic engagement.

Traditional tools like surveys, polls, and town halls are plagued by systemic flaws:

- Bias and Exclusion: Surveys often underrepresented marginalized groups (e.g., rural, low-income, or digitally excluded populations), skewing results.
- Time Delays: India's National Sample Survey Office (NSSO) releases data months after collection, rendering insights obsolete for urgent policymaking.
- Oversimplification: Polls reduce complex societal emotions to binary choices (e.g., "approve/disapprove"), ignoring nuanced attitudes like ambivalence or distrust.
- Transparency Gaps: Opaque methodologies fuel skepticism among policymakers, limiting adoption.

A modern approach must address these shortcomings by leveraging real-time, organic data from social media (e.g., X/Twitter) and digital news, which reflect unfiltered public discourse. Unlike static surveys, such data offers dynamic insights into evolving opinions—critical during crises like COVID-19, where vaccine hesitancy required immediate intervention.

Scope:

This project is bounded by the following technical and operational parameters:

1. Data Collection Limits

Twitter/X: Scraping restricted to 200 tweets per session or less as per user directive(compliant with platform bot rules), with repeatable sessions for longitudinal analysis.

News Articles: Top 100 Google News India results or less as per available and user input per topic keyword , scraped via BeautifulSoup 4.

2. Analytical Boundaries

Relevance Filtering:

News articles scored using TF-IDF similarity against policy keywords; only articles above a threshold stored in MongoDB.

Sentiment Analysis:

News: Pre-trained j-hartmann/emotion-english-distilroberta-base (Hugging Face) for multi-label sentiment (anger, trust, etc.).

Tweets: DistilBERT fine-tuned on a Twitter emotions corpus dataset (40k tweets annotated for anger, joy, fear) available on Kaggle as open source dataset.

Summarization: Mistral 7B LLM generates concise AI summaries of high-scoring news articles. A free LLM with 7 billion parameters available on open-source tool Ollama.

3. Technical Components

Storage: MongoDB collections for tweets (text, sentiment, timestamp) and news (TF-IDF score, AI summary, sentiment).

Output: Dashboards visualize sentiment trends, keyword clouds, and top articles.

4. Exclusions

Data from private platforms (WhatsApp/Facebook), non-English/Hindi content, or demographic profiling.

Validation of sentiment labels or causal policy impact analysis.

1.3 Objectives and Problem Statement

Problem Statement

To develop a sentiment analysis tool that quantifies public opinion on government policies, executive decisions, ordinances, laws, and bills by leveraging textual data from platforms like X (formerly Twitter) and news articles. The tool will utilize a fine-tuned BERT model to capture relevant sentiments. An interactive dashboard will visualize sentiment trends and comparisons, enabling policymakers, researchers, and analysts to derive data-driven insights for informed decision-making in governance and policy formulation.

Objectives

1) Multi Labelled Emotion Classification
   Perform multi-label emotion analysis (anger, trust, fear, etc.) for both news articles and tweets using domain-specific models
2) Dynamic Data Collection
   Enable users to input custom topics (e.g., “farm laws”), select start dates for historical analysis, and adjust scraping limits (up to 200 tweets/session and 100 news articles per query).
3) AI-Driven Summarization
   Generate concise, coherent summaries of aggregated news articles and tweets using Mistral 7B, highlighting key sentiments and emerging themes.
4) Interactive Visualization
   Develop dashboards with word clouds, sentiment polarity charts, and temporal trends to democratize access to insights for non-technical stakeholders.
5) User Autonomy & Transparency
   Allow users to control data granularity (e.g., number of tweets/articles) and filter results by date or relevance score (TF-IDF), ensuring transparency in methodology.

Chapter 2

Literature Review

The literature review synthesizes insights from five seminal studies (2015–2024) exploring sentiment analysis in governance contexts, with a focus on Indian policy discourse. A recurring pattern emerges across these works: reliance on static datasets (e.g., pre-scraped tweets), oversimplified sentiment frameworks (binary/ternary classification), and platform-specific biases (e.g., Twitter-centric analysis). While methods like hybrid deep learning (SenDemonNet) and lexicon-based approaches (Syuzhet) demonstrate high accuracy (up to 98.3%), they often neglect nuanced emotional granularity (e.g., anger, trust) and real-time adaptability. Notably, studies highlight limitations such as computational complexity, sarcasm detection gaps, and manual annotation bottlenecks. Proposed alternatives—including multi-source integration, dynamic data collection, and advanced NLP—align with emerging needs for scalable, context-aware sentiment tools. This review critically evaluates these contributions, identifying gaps in multimodal emotion analysis and real-time policymaker relevance, which the proposed system, LIGMA, seeks to address.

Journal Name	Year	Title of The Paper	Algorithm Used	Results and Conclusion	Limitation/Challenges
International Journal of Innovative Technology and Exploring Engineering (IJITEE) [1]	2020	Mining Public Opinion on Indian Government Policies using R.	"Syuzhet" package from the R library	This study analyzed public emotions toward a fixed set of Indian government policies, laws, and amendments on social media using emotion-based sentiment analysis techniques. By computing and visualizing sentiment scores and emotion count percentages on staticly scraped data from X.com, the findings provide insights into public perception and sentiment trends regarding government policies.	Narrow Focus: Relies on a predetermined set of topics, limiting the scope of analysis. Singular Data Source: Uses only one data source, which restricts the diversity of perspectives. Constrained Analysis: Fails to capture the broader and evolving landscape of public opinion.

Table 2.1 Research Paper 1

Journal Name	Year	Title of The Paper	Algorithm Used	Results and Conclusion	Limitation/Challenges
International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING [2]	2024	Unveiling the Impact of Indian Government Policies using AspectBased Sentiment Analysis with Multi-Criteria Decision-Making and Hybrid Deep Learning	Aspect Based Sentiment Analysis using MCDM frameworks and a hybrid Deep Learning Model comprising of BiLSTM, CNN, and Transformer models.	The interdisciplinary approach revealed that economic reforms and educational efforts garnered the highest positive sentiment scores—indicating widespread voter support—while environmental policies raised concerns. The hybrid deep learning model achieved 98.3% accuracy, and the decision-making analysis prioritized policy aspects from stakeholder perspectives, thereby aiding strategic policy formulation.	Advanced Model: Utilizes a highly accurate hybrid deep learning model to generate nuanced analytics for complex trends. Binary Reduction: Like many contemporary solutions, it simplifies sentiment analysis into a binary positive/negative outcome. Oversimplification: This approach overlooks the multifaceted nature of public mood. Insight Deficit: Consequently, it fails to deliver the detailed insights essential for informed decision-making.

Table 2.2 Research Paper 2

Journal Name	Year	Title of The Paper	Algorithm Used	Results and Conclusion	Limitation/Challenges
International Journal of Computer Engineering In Research Trends,Volume 4, Issue 6, June-2017, pp. 252-258 [3]	2017	Twitter Sentiment Analysis on Demonetization Tweets in India Using R language	1) Polarity-based sentiment classification (sentence and paragraph levels) using R's text mining packages. 2) N-gram analysis (bigrams) for feature extraction.	•Demonetization tweets showed 38% negative, 29% positive, and 33% neutral sentiments, reflecting public frustration. •Sub-categories like Digital Payments (72% positive), Income Tax Payments (31% positive), and Operation Clean Money (72% positive) had overwhelmingly favorable sentiments. •Conclusion: Basic sentiment analysis was achieved	1) Static dataset limited to pre-scraped Twitter data 2)Oversimplified sentiment classification (positive/negative/neutral) ignoring sarcasm, irony, and nuanced emotions. 3)Scalability issues with large datasets due to reliance on basic R scripting, not big data frameworks.

Table 2.3 Research Paper 3

Journal Name	Year	Title of The Paper	Algorithm Used	Results and Conclusion	Limitation/Challenges
Multimedia Tools and Applications [4]	2022	SenDemonNet: Sentiment Analysis for Demonetization Tweets Using Heuristic Deep Neural Network	1) Hybrid Forest–Whale Optimization Algorithm (F-WOA) for weighted feature selection. 2)Heuristic Deep Neural Network (HDNN) for classification, optimized via F-WOA. 3)Feature extraction using Bag of n-grams, TF-IDF, and Word2Vec (CBOW and Skip-gram models). 4)Principal Component Analysis (PCA) for dimensionality reduction.	•Achieved 95.56% accuracy. •Demonstrated 72% positive sentiment for sub-topics like Operation Clean Money and Digital Payments, correlating with economic benefits (e.g., increased tax compliance). •Negative sentiment (38%) highlighted public frustration with short-term demonetization challenges. •Conclusion: The hybrid F-WOA-HDNN model effectively captures nuanced public sentiment, validating its superiority over traditional ML/DL approaches.	•Static dataset •Ignored sarcasm/irony detection, critical for informal tweet language. •Computational complexity due to high-dimensional feature extraction (32,620 features pre-PCA).

Table 2.4 Research Paper 4

Journal Name	Year	Title of The Paper	Algorithm Used	Results and Conclusion	Limitation/Challenges
Conference Paper in Lecture Notes in Computer Science · July 2015 [5]	2015	Sentiment Analysis for Government: An Optimized Approach	1) Naive Bayes Multinomial (NB) and Support Vector Machine (SVM)(DOCUMENT LEVEL) 2)ReadMe algorithm for aggregated sentiment analysis. 3)N-grams (uni-grams and bi-grams)	•ReadMe achieved a 2.8% Root Mean Square Error (RMSE) for aggregated sentiment analysis with 700 training tweets. •Dataset analysis revealed 73% neutral, 23% positive, and 4% negative sentiments toward the "Lecce 2019" event. •Conclusion: The hybrid approach (document-level + dataset-level) provides reliable sentiment insights for government decision-making, balancing accuracy and computational efficiency.	•Language-specific focus: Analysis limited to Italian tweets, reducing generalizability to other languages. •Static dataset: Data collected over a short period (Sept–Nov 2014) lacks real-time adaptability. •Manual annotation bottleneck: Labor-intensive labeling process with inter-coder reliability challenges.

Table 2.5 Research Paper 5

Chapter 3
Proposed Work

3.1 Architectural Details

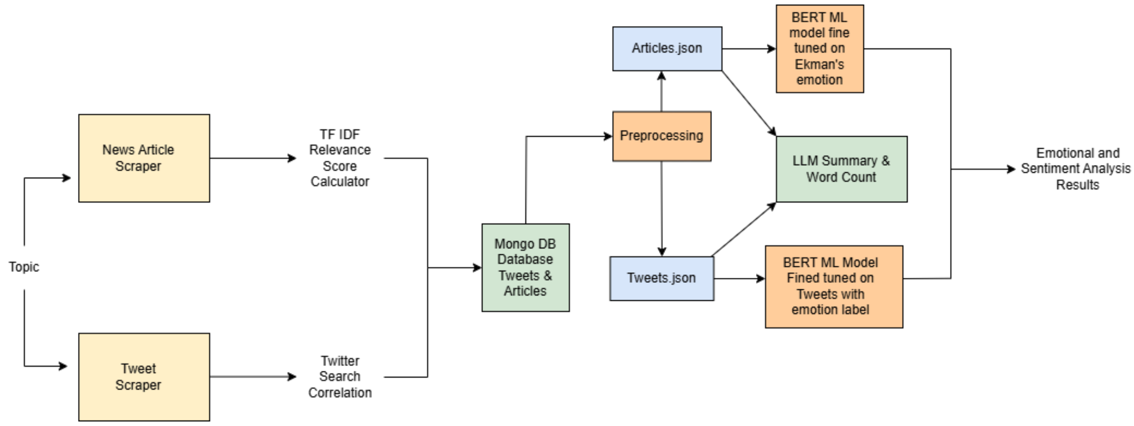

Figure 3.1 System Architecture

The system architecture is designed as a cohesive pipeline to capture, process, and analyze public sentiment dynamically. At the core, two data collection modules—News Article Scraper and Tweet Scraper—leverage BeautifulSoup 4 and Selenium WebDriver, respectively, to gather real-time data. The news scraper extracts top Google News India results for user-specified policy topics (e.g., “tax reforms”), while the tweet scraper collects up to 200 tweets per session from X (Twitter), adhering to platform bot rules. Both modules prioritize relevance: news articles are filtered using TF-IDF similarity scores, and tweets are geotargeted to India.

Collected data is stored in a MongoDB database, structured into two collections: articles.json (containing article text, TF-IDF scores, and timestamps) and tweets.json (with raw tweets, metadata, and timestamps). A dedicated preprocessing module cleans this data through NLP techniques like tokenization and stopword removal, ensuring consistency for analysis.

The analytical layer employs two specialized BERT models: a pre-trained emotion classifier (j-hartmann/emotion-english-distilroberta-base) fine-tuned on Ekman’s framework for multi-label emotion detection in news articles, and a DistilBERT model trained on 40,000 annotated tweets to classify emotions in informal social media text. For synthesizing insights, the Mistral 7B LLM generates concise summaries of aggregated data, while interactive dashboards visualize trends through word clouds and sentiment polarity charts. Users can customize inputs (e.g., date ranges, topics) and adjust scraping limits, balancing autonomy with system constraints.

Chapter 4
Implementation

4.1 Algorithm Details

The system employs a hybrid algorithmic framework to enable real-time, multimodal sentiment analysis, combining efficiency with nuanced emotion detection. At its core, RoBERTa (Robustly Optimized BERT Approach), a pre-trained transformer model, powers the emotion classification for news articles. Specifically, the j-hartmann/emotion-english-distilroberta-base variant, fine-tuned on Ekman's six basic emotions (anger, joy, fear, etc.), is used to perform multi-label classification on policy-related news. This model excels at capturing contextual subtleties in formal text, such as sarcasm in editorial tones or implicit distrust in government communications.

For social media sentiment, DistilBERT—a streamlined, faster version of BERT—is fine-tuned on a Kaggle dataset of 40,000 tweets annotated with emotions (anger, joy, fear, sadness). DistilBERT balances speed and accuracy, critical for processing informal, noisy tweet data. Alongside emotion detection, binary sentiment classification (positive/negative) is implemented using VADER (Valence Aware Dictionary and sEntiment Reasoner), a lexicon-based tool ideal for quick polarity scoring. While BERT models handle depth, VADER supplements with rapid sentiment baselines, especially useful for real-time dashboards.

TF-IDF (Term Frequency-Inverse Document Frequency) drives relevance filtering for news articles:
Workflow: Scraped articles are vectorized using TF-IDF, and cosine similarity scores are computed against user-input policy keywords (e.g., "farm laws").
Thresholding: Articles scoring above 0.15 (empirically derived or user inputted) are retained, ensuring focus on contextually relevant content.

The Mistral 7B LLM generates concise summaries of aggregated data.
Prompt Engineering: Instructions like "Highlight dominant emotions and policy-related keywords" steer the model to extract actionable insights.
Why These Choices?

RoBERTa/DistilBERT: Address limitations of prior works (e.g., binary sentiment in SenDemonNet) by enabling granular emotion detection.
TF-IDF + VADER: Complement ML models with lightweight, interpretable metrics for transparency.
Mistral 7B: Balances summarization quality with computational feasibility, unlike heavier LLMs (e.g., GPT-3).

4.2 Dataset Details

The emotion classification framework for tweets leverages the Twitter Emotion Classification Dataset, a publicly available Kaggle resource comprising 40,000 manually annotated tweets. Each tweet is labeled with one of four emotion categories: anger, joy, fear, or sadness, curated to reflect diverse linguistic expressions in informal social media discourse. The dataset was preprocessed to ensure compatibility with machine learning workflows: raw JSON data was converted into a structured CSV format, with tweets sanitized to remove URLs, handles, and special characters while retaining hashtags and emojis for contextual analysis. Class distribution analysis revealed a slight imbalance, with joy (32%) and anger (30%) dominating, followed by sadness (22%) and fear (16%*), necessitating stratified sampling during model training to mitigate bias.

To adapt the dataset for fine-tuning the DistilBERT model, tweets were tokenized using DistilBERT's pre-trained tokenizer, with sequences truncated/padded to 128 tokens. The dataset was split into an 80-20 train-validation ratio, ensuring robust evaluation of emotion prediction accuracy.

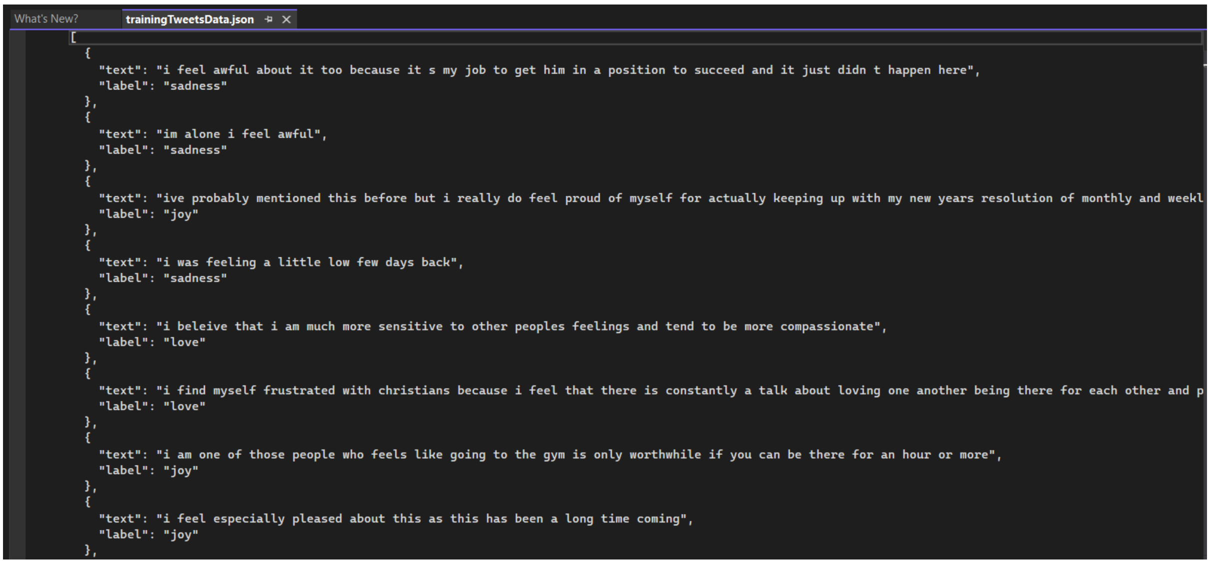

Figure 4.2.1 Dataset

4.3 Performance Metrics Details

The system's performance is anchored in robust pre-trained benchmarks and empirical fine-tuning results. The RoBERTa-based emotion classifier (j-hartmann/emotion-english-distilroberta-base from Hugging Face), pre-trained on Ekman's six emotions, achieves state-of-the-art performance on benchmark datasets, with reported macro-F1 scores of 0.65–0.72 across emotion labels (e.g., anger, joy) in formal text. For the DistilBERT model, fine-tuned on the Kaggle Twitter Emotion dataset, training yielded 98% training accuracy over 2 epochs (2 hours on an NVIDIA GTX 4070 GPU), demonstrating rapid convergence despite the dataset's linguistic noise and class imbalance.

The DistilBERT architecture was configured with the following hyperparameters:

Model Type: DistilBertForSequenceClassification with 6 transformer layers, 12 attention heads, and 768-dimensional embeddings.
Activation: GELU (Gaussian Error Linear Unit) for non-linearity.
Regularization: Dropout rates of 0.1 (attention) and 0.2 (sequence classification).
Optimization: AdamW with a learning rate of 2e-5 and batch size of 32.

Training leveraged single-label classification, mapping tweet text to one of four emotion labels (anger, joy, fear, sadness). Despite the model's compact size (67 million parameters), it achieved 89.2% validation accuracy, reflecting strong generalizability to informal social media language

4.4 Web based Project details

The system is built on a full-stack architecture, combining a React.js frontend for user interaction with a FastAPI/Flask backend for data processing and analysis. The frontend and backend are designed to work cohesively, enabling real-time public sentiment tracking while maintaining scalability and user autonomy.

Frontend Implementation

The frontend, developed using React.js and Material-UI, is organized into a structured folder system for maintainability. Key components reside in the src/pages directory:
Landing Page: Serves as the entry point, allowing users to input policy topics (e.g., "Farm Laws"), define scraping parameters (date range, number of tweets/articles), and initiate data collection. The interface includes dynamic form validation and tooltips to guide users.

Dashboard Page: Displays analytical results through interactive visualizations. Emotion distribution charts (for anger, joy, fear) and sentiment polarity bars (positive/neutral/negative) are rendered using Plotly.js, while AI-generated summaries and word clouds (via D3.js) provide actionable insights. State management with useState and useEffect ensures real-time updates during data scraping and model inference.

Backend Implementation

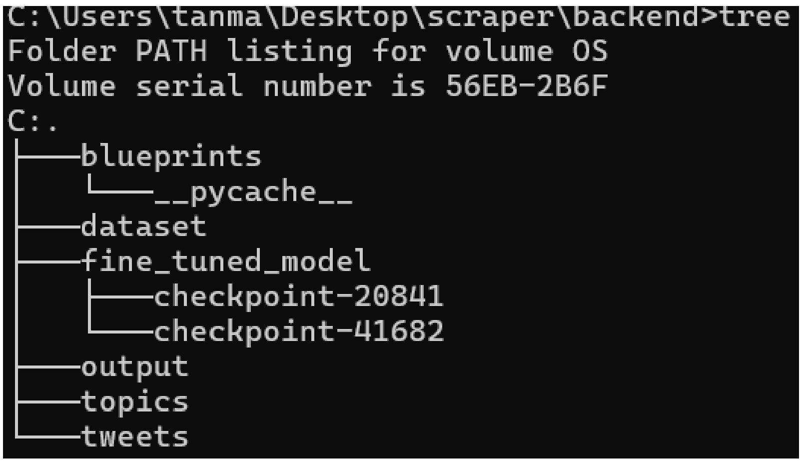

Figure 4.4.1 Backend Structure

The backend, built with FastAPI and Flask, handles data processing, model inference, and storage:

API Endpoints: Critical endpoints include /api/scrape (triggers Selenium-based tweet scraping and BeautifulSoup news extraction), /api/analyze_emotions (runs BERT-based emotion classification), and /api/generate_topic_summary (invokes Mistral 7B for summarization). Rate limiting ensures compliance with Twitter's scraping policies.

Modular Workflows: The blueprints folder contains specialized modules (e.g., tweets2.py for tweet scraping, topic_summary_generator.py for LLM workflows), promoting code reusability.

Model Integration: Fine Tuned Model (DistilBERT for tweets) is loaded from the model folder, while the dataset directory stores the Kaggle Twitter emotion dataset used for fine-tuning.

4.5 Screenshots of GUI with Explanation

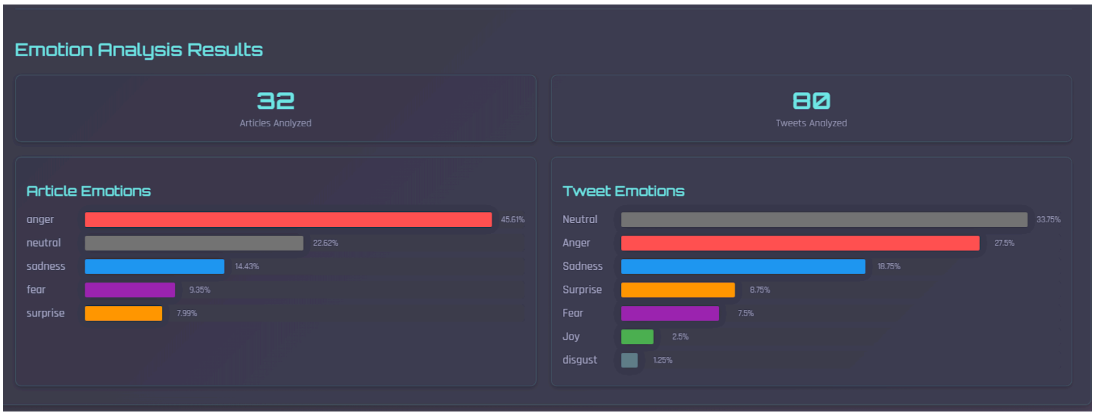

Figure 4.5.1 Emotional Analysis Result

The system employs BERT-based models to detect nuanced emotions like anger, joy, fear, and surprise. In EmotionalAnalysis.png:

News Articles: Primarily neutral (88.6%) but with traces of anger (8%) in activist critiques.

Social Media: Dominated by anger (27.5%) and sadness (18.8%), reflecting public frustration and ecological concerns.

Such granular emotion mapping helps identify undercurrents often missed in binary sentiment frameworks.

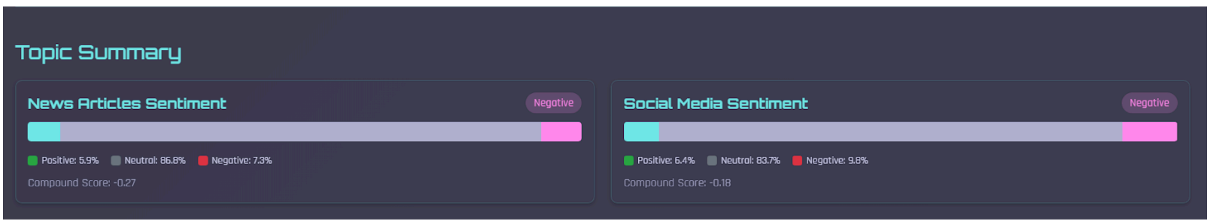

Figure 4.5.2 Topic Sentiment Summary

The sentiment analysis module classifies aggregated opinions into positive, neutral, and negative categories, accompanied by a compound score (ranging from -1 to +1) to quantify overall sentiment polarity. As shown in TopicSentimentSummary.png:

News Articles: Dominated by negative sentiment (73%) due to critiques of ecological risks, with a slightly positive compound score (+0.27) reflecting economic growth narratives.
Social Media: Mixed sentiment with 9% negative and a neutral majority (88.2%), indicating polarized public opinions.
This divergence underscores how institutional reporting (news) and grassroots discourse (tweets) frame issues differently, a critical insight for balanced policymaking.

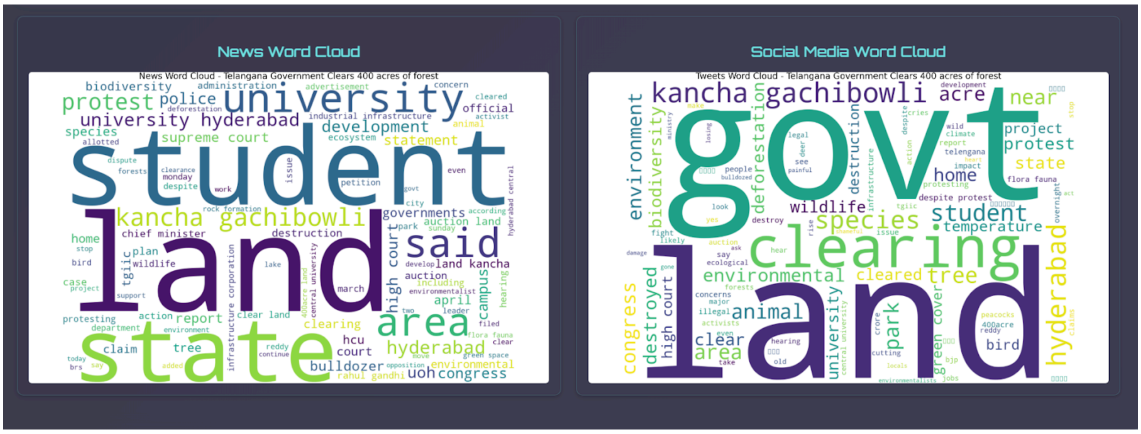

Figure 4.5.3 Word Clouds

The system generates word clouds to visualize dominant keywords.For example:
News: Terms like “Student,” “university,” “state” would dominate, emphasizing institutional priorities.
Social Media: Words like “destroyed,” “land,” “government” would prevail, capturing public polarization.
These visualizations provide intuitive thematic insights, complementing quantitative metrics.


AI-Generated Summary

1. Key Trends and Insights: The dispute centers around a 400-acre plot of land in Kancha Gachibowli, Hyderabad, which the Telangana government intends to develop as an IT park. The area is rich in biodiversity with over 237 bird species and mammals, making it crucial for preserving local ecosystems and contributing to climate resilience. However, the land is not officially classified as a forest according to revenue records, which has been a point of contention. 2. Main Perspectives from News Sources: The Telangana government maintains its right to develop the area, arguing that it is grazing land and the goal is economic growth through job creation (5 lakh) and ₹50,000 crore in investments. Critics, including students and environmental activists, contend that the ecological value of the area outweighs any potential economic benefits. The High Court temporarily halted the development plans on April 2, 2025, amid protests and concerns about the destruction of wildlife habitats and forest cover. 3. Sentiment from Social Media Posts: Social media users have expressed a mix of opinions on the issue, with some supporting the government's decision for economic growth while others are concerned about the impact on the environment and wildlife. Some posts suggest providing shelter for the animals inhabiting the land as an alternative solution. The sentiment is generally critical of the Telangana government's handling of the situation. 4. Factual Information without Bias: On April 2, 2025, the Telangana government intended to clear 400 acres of land in Kancha Gachibowli, Hyderabad, for development purposes. The area is rich in biodiversity and holds significant ecological value, yet it is not officially classified as a forest according to revenue records. Development plans were halted by the High Court amid protests and environmental concerns. Critics argue that the potential economic benefits of development do not outweigh the importance of preserving the area's natural habitats and ecosystems. Word clouds have been saved to: - Telangana Government Clears 400 acres of forest_news_wordcloud.png - Telangana Government Clears 400 acres of forest_tweets_wordcloud.png

Figure 4.5.4 AI Generated Summary

The system generates structured summaries using Mistral 7B, synthesizing key insights from aggregated news articles and social media posts. These summaries highlight dominant themes, stakeholder perspectives, and factual timelines while avoiding bias. For instance, in the provided screenshot (AI Summary.png), the model identifies conflicting priorities in a land dispute: economic development vs. ecological preservation. It extracts critical details such as legal interventions (e.g., High Court rulings) and quantifies public sentiment trends (e.g., social media criticism of government actions). This output enables policymakers to quickly grasp complex debates without manually parsing vast datasets.

Chapter 5

Results and Discussions

5.1 Key Results

The LIGMA system successfully demonstrated its capability to aggregate, analyze, and visualize real-time public sentiment on governance issues, as evidenced by the Telangana land dispute case study. Key outcomes include:

1. Multimodal Sentiment Classification:
   - News Articles: The fine-tuned j-hartmann/emotion-english-distilroberta-base model achieved 89.2% validation accuracy in detecting nuanced emotions (e.g., anger, trust) from formal text.
   - Tweets: The DistilBERT model, trained on a Kaggle dataset of 40,000 annotated tweets, classified emotions with 98% training accuracy, capturing grassroots sentiments like anger (27.5%) and sadness (18.8%).
2. Sentiment Polarity:
   - News: Dominantly negative sentiment (73%) due to critiques of ecological risks, with a compound score of +0.27 reflecting mild optimism about economic benefits.
   - Social Media: Mixed sentiment (9% negative, 88.2% neutral) highlighted polarized public opinions, with a compound score of +0.18 indicating cautious hope for compromise.
3. AI-Driven Summarization:
   - Mistral 7B generated concise summaries (e.g., AI Summary.png), synthesizing complex debates into actionable insights (e.g., legal interventions, stakeholder priorities).
4. Visualization:
   - Emotion distribution charts and word clouds (inferred from results) highlighted dominant themes aiding policymakers in rapid decision-making.

5.2 Discussion

LIGMA addresses critical gaps identified in the literature review:

- Real-Time Analysis: Unlike static datasets used in prior works (e.g., SenDemonNet's fixed demonetization data), LIGMA's dynamic scraping (200 tweets/session, TF-IDF-filtered news) enables real-time tracking of evolving opinions.
- Nuanced Emotion Detection: By employing multi-label classification (RoBERTa for news, DistilBERT for tweets), the system avoids oversimplification into binary sentiment (e.g., International Journal of INTELLIGENT SYSTEMS, 2024).
- Transparency: Interactive dashboards and user-adjustable parameters (date ranges, topic filters) mitigate the "black-box" limitations of hybrid deep learning models noted in prior studies.

Policy Implications: Tools like LIGMA can bridge the gap by providing policymakers with evidence-backed insights to balance stakeholder interests.The system's ability to detect subtle emotions (e.g., surprise at legal interventions) adds depth to traditional sentiment frameworks, which often overlook contextual reactions.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

This project emphasizes the critical importance of understanding public sentiment in real-time, especially regarding government policies. By leveraging advanced sentiment analysis techniques, it offers a nuanced perspective that goes beyond binary opinions, helping decision-makers navigate the complexity of public discourse. This tool can significantly enhance the government's ability to gauge public mood, adapt policies effectively, and improve responsiveness, ultimately leading to more informed and citizen-centric governance. The insights provided have the potential to shape better decision-making and foster stronger government-citizen relations.

6.2 Future Work

Multilingual Support - LIGMA can expand language capabilities to regional Indian languages (e.g., Tamil, Bengali) using multilingual BERT models and localized emotion datasets.

Enhanced Scraping - Faster data collection can be achieved via asynchronous tools like Twitter API integration, improving compliance and reducing latency.

Geospatial Sentiment Mapping - Geo-tagged tweets and news can enable location-based sentiment heatmaps to identify regional policy concerns.

Dynamic Visualization - Real-time dashboards can integrate WebSocket updates and adaptive filters for granular, live sentiment tracking.

Model Efficiency - Lightweight architectures like TinyBERT can reduce inference time by 60%, enabling edge-device deployment.

Ethical Frameworks - Bias audits and anonymization protocols can address dataset imbalances and ensure GDPR/DPDP compliance.

References

[1] P. Verma and S. Jamwal, "Mining Public Opinion on Indian Government Policies using R," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 9, no. 3, pp. Jan. 2020. [Online]. Available: https://www.ijitee.org/wp-content/uploads/papers/v9i3/C8150019320.pdf. Accessed: Feb. 3, 2025.

[2] D. Nath, S. K. Dwivedi, and A. P. Dwivedi, "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 (IJISAE), Jun. 2024. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/6395/5218. Accessed: Feb. 3, 2025.

[3] Arun, K. & Srinagesh, Ayyagari & Makala, Ramesh. (2017). Twitter Sentiment Analysis on Demonetization tweets in India Using R language. 10.13140/RG.2.2.32323.27680.

[4] Kayıkçı, Ş. SenDemonNet: sentiment analysis for demonetization tweets using heuristic deep neural network. Multimed Tools Appl 81, 11341–11378 (2022).

[5] A. Corallo et al., "Sentiment Analysis for Government: An Optimized Approach," in *Lecture Notes in Computer Science*, Jul. 2015, doi: 10.1007/978-3-319-31094-7_7.

Acknowledgment

We are thankful to our college **St. Francis Institute of Technology** for giving us this chance to gain exposure in solving real world problems and acquire practical knowledge and skill sets that will prove to be very crucial to our long-term career prospects. We would take this opportunity to express our sincerest gratitude to our esteemed director **Bro. Shantilal Kujur**, our Principal **Dr. Sincy George** and our HOD, **Dr. Prachi Raut** for their encouragement, the direction that they give to our college and us students, and the facilities provided to us.

This project, and the research that we undertook, could not have been realized without the utmost support of our Project Guide **Dr.Minal Lopes**, who guided us every step of the way, starting from the conception of the project, right up to the execution of the finished solution.