# Sentiment Tracker for Indian Government Schemes Using Twitter

*Authors : Sujal , Rishi Kumar Jakhmola , Mohammad Anas , Rahul Kumar*

**IMS Ghaziabad University Courses Campus**

## ABSTRACT

Alright, let's get real—government welfare schemes in India are basically the country's attempt to patch up some big, gnarly problems: think housing shortages, tricky healthcare access, spotty digital stuff, and, oh yeah, making it easier for folks to kickstart their own businesses. But here's the kicker—most of these programs chug along without any decent way to actually hear what people think, you know, as the policies roll out and (hopefully) get better.

So, what did we do? We built this sentiment tracker that grabs tweets about four big-shot government schemes—PM Awas Yojana, Ayushman Bharat, Digital India, and Startup India—right from Twitter (thanks, Tweepy and API v2, you legends), for the chunk between January 2024 and May 2025. Cleaned up the tweets (gotta get rid of all the junk), chopped 'em up, lemmatized—yeah, that's a real word—and then let TextBlob loose to slap a sentiment score on each: positive, negative, or meh (neutral, for the fancy folks).

Total tweets analyzed? 8,200. Not too shabby. Then we made it all look good with a Streamlit dashboard—word clouds, charts, those timey-wimey plots that show how opinions shift. What stood out? PM Awas Yojana and Ayushman Bharat were the stars—people mostly dug 'em. Startup India? Eh, not so much—lotta grumbling about money hassles and red tape. Digital India? Kind of a mixed bag. Folks are hyped about better internet, but the infrastructure struggles still get people riled up.

Long story short: tracking social media sentiment like this is more than just geeky fun. It actually gives policymakers a chance to tweak things in real time—finally, some hope for government stuff that actually listens.

**KEYWORDS:** Sentiment Analysis, Indian Government Schemes, Twitter, TextBlob, Streamlit, NLP, CivicTech.

## INTRODUCTION

Government schemes like PM Awas Yojana, Ayushman Bharat, Digital India, and Startup India are more than just policies—they are promises. Promises of a better life, a safer home, access to quality healthcare, digital empowerment, and the ability to dream big through entrepreneurship. Launched with great ambition and backed by significant investment, these schemes reflect India's commitment to inclusive development.

But beyond launch speeches and budget allocations, one crucial question remains: What do the people think?

Public perception plays a huge role in determining whether a government initiative is successful or not. If citizens believe in a scheme—if they feel seen, supported, and heard—they are more likely to engage with it. If they don't, they might ignore it, criticize it, or worse, lose trust in public institutions altogether.

This is where social media, and in particular Twitter, comes in. Every day, thousands of Indians take to Twitter to voice their opinions. Some share positive experiences—like receiving their first subsidized home under PMAY or getting cashless treatment through Ayushman Bharat. Others point out flaws, such as application hurdles, bureaucratic red tape, or poor digital infrastructure. These opinions, shared in real time, form a living pulse of public sentiment.

In today's world, understanding this pulse is critical. Traditional surveys and paper-based feedback mechanisms are too slow and often limited in scale. But Twitter, with its fast-moving content, trending hashtags, and open

conversations, offers something new: a window into real-time public thinking, available to anyone with the tools to make sense of it.

**Why Twitter?**

With over 20 million active users in India, Twitter is now more than a social platform—it's a public square. Politicians, journalists, activists, and everyday citizens use it to talk about everything from cricket and cinema to health policy and education. Major government announcements trend within minutes. News breaks here before it hits television. Public outrage or celebration finds its voice in tweets.

The beauty of Twitter lies in its simplicity. A single tweet—just 280 characters—can carry immense emotional weight. People don't hold back. They tweet about delays in government services, their struggles to access benefits, their happiness on receiving help, or even their disbelief in bureaucratic promises. And because of the platform's design—its hashtags, mentions, and replies—these emotions travel fast, gaining likes, retweets, and visibility.

This makes Twitter a goldmine for sentiment analysis—if we can decode what's being said, how it's being said, and why.

**The Need for a Sentiment Tracker -**

Think of this study as building a weather radar, but instead of tracking rain clouds or temperature shifts, we're tracking public emotion. Our goal is to build a simple, insightful system that can tell us how people are feeling about specific government schemes—whether they're hopeful, frustrated, confused, or proud.

Let's say the government announces an increase in PMAY funding in the next budget. On the surface, this might look like good news. But what if a surge of tweets expresses skepticism? What if people are saying, "Nice announcement, but I've been waiting for my house for three years"? These tweets reveal something that numbers alone cannot—the lived experience of policy.

Or take Digital India. While one set of user praises increased digital access, another set might be tweeting from rural areas complaining about slow internet, broken portals, or non-functioning apps. A sentiment tracker can highlight these voices, helping policymakers make data-informed adjustments.

**What We're Doing ?**

This study builds a Twitter-based sentiment tracking system that looks at how people are talking about four flagship Indian government schemes: PM Awas Yojana, Ayushman Bharat, Digital India, and Startup India.

We started by gathering tweets that mention these schemes using relevant hashtags and keywords (like #PMAY, "Startup India," or "Ayushman Bharat card"). Using the Twitter API (and the Tweepy Python library), we collected over 8,000 tweets shared between January 2024 and May 2025.

Next, we cleaned the data. Tweets are often messy—they include emojis, slang, hashtags, links, and even sarcasm. So we preprocessed the text: converting it to lowercase, removing irrelevant elements like URLs and mentions, filtering out stop words (common words like "the" and "is"), and breaking down the sentences into meaningful tokens (using SpaCy for both English and Hindi).

To detect sentiment, we used a tool called TextBlob. It scores each tweet on a scale from -1 to +1—where negative numbers represent negative sentiment, positive ones represent positive sentiment, and scores near zero are neutral. We set simple thresholds to classify the tweets accordingly.

Then, we built an interactive dashboard using Streamlit, where users can view:

Word clouds that show commonly used terms for each scheme

Sentiment distribution graphs to see how opinions are spread

Time-series plots to observe how sentiment changes over days and months

This dashboard is lightweight, open-source, and accessible—not just to researchers but to policymakers, journalists, or even curious citizens.

**What We're Asking ?**
Our study is centered around three key questions:
What are people saying about each scheme?
Are citizens praising affordability? Complaining about red tape? Sharing success stories?
How does sentiment shift over time?
Do major events—like budget speeches, scheme relaunches, or viral videos—trigger changes in how people feel?
What themes keep popping up?
Are certain keywords or concerns (like "delay", "access", "benefits") repeatedly mentioned across tweets?
Why It Matters
This is more than an academic exercise. Real-time sentiment tracking has the potential to make government more responsive.

Imagine if, within 24 hours of a new announcement, policymakers could get a pulse on whether citizens feel hopeful or skeptical. Imagine being able to identify that most complaints about Startup India in April 2025 came from first-time entrepreneurs in tier-2 cities who couldn't access funding.

**The applications are wide:**

- Policymakers can tweak policies based on real feedback
- Government departments can identify communication gaps
- Researchers can study how sentiment shapes policy outcomes
- Journalists and activists can surface ignored public concerns
- Most importantly, citizens can be heard—not just during elections but every day.

**The Bigger Picture -**

Globally, sentiment analysis is being used in elections, brand reputation management, crisis response, and more. In India, it's still an emerging field in the policy space. This study contributes to the growing movement of civic tech—where technology is used to strengthen democracy, transparency, and participation.

**Our tracker is simple by design, but its implications are powerful. In the future, we envision building on this work by:**

Using advanced AI models (like Indic-BERT) to handle sarcasm and regional language better

Adding sentiment from other platforms like YouTube, Koo, or Reddit

Expanding to include more schemes or state-level programs

## MATERIAL AND METHODS

Alright, let's break it down real-world style—no stiff academic jargon or robotic sentences. Here's how the whole operation went down:

So, the mission? Build a chill, snappy dashboard that grabs tweets about Indian government schemes, figures out if folks are loving or hating them, and then slaps all that data on some eye-candy visuals. All in real time, too. No waiting around.

**Tools & Tech (aka, What We Used to Make the Magic Happen)**

Python was the main player here—because honestly, what can't Python do? It's got, like, a million data science tools and actually makes sense to use.

- Twitter API (via Tweepy): Basically the fishing rod for reeling in tweets, live, using whatever keywords we wanted (stuff like "Ayushman Bharat" or "Digital India").

- **TextBlob**: This little library does the heavy lifting for sentiment analysis. It figures out if a tweet is sunshine and rainbows or a storm cloud of rage.

- **Streamlit**: Makes dashboards look good without you needing to be a front-end wizard.

- **Matplotlib & Seaborn:** For whipping up charts and graphs, because everybody loves a good visual.

- **WordCloud**: Turns a pile of words into those funky blobs where the popular ones jump out at you.

- **Pandas**: The bread and butter for crunching and reshaping data tables.

- **Pytrends**: Hooks into Google Trends so we could see what's hot and what's not outside of Twitter.

**Grabbing the Data**

First, you gotta get through Twitter's security—grab some auth tokens, so you're not just banging on the door. For each scheme (think PM Awas Yojana, Startup India, etc.), we tossed in some search terms and pulled tweets in English using Tweepy's search_recent_tweets(). You could pick your batch size—want 10 tweets? 100? Up to you.

**Cleaning the Mess**

Tweets are messy. Like, seriously. Mentions, hashtags, weird links, emojis, the whole circus. We cleaned all that junk out with regular expressions, dropped everything to lowercase, and ditched the punctuation. Gotta start with clean data if you want the analysis to mean anything.

**Sentiment Analysis—The Fun Bit**

TextBlob took the cleaned tweets and rated their "vibe" on a scale from -1 (major yikes) to +1 (all good). Anything above zero? Positive. Below zero? Negative. Exactly zero? Meh, neutral. Slapped those labels into a new DataFrame column for easy charting later.

**Making It Look Good—Visuals**

Once everything had a sentiment, we made it pop:

- **Bar charts:** Showed how many tweets were positive, negative, or just chilling in the middle.

- **Line graphs:** Let us see how moods shifted over time.

- **Word clouds:** All the hot words getting thrown around, in big, bold letters.

- **Google Trends**: Used Pytrends to track what people were searching for, outside of Twitter, over time.

**The User Interface (aka, Where the Magic Happens)**

Everything ended up in a slick, idiot-proof Streamlit web app. Click a scheme, choose how many tweets you want, and boom—real-time results. You don't need to know how to code (or even what a DataFrame is) to use it. Perfect for classrooms, presentations, or just curious folks poking around.

And yeah, that's the whole toolkit and workflow—no fancy suits required.

---

## LITERATURE REVIEW

Sentiment analysis is extensively employed to measure public sentiment over topics ranging from politics, marketing, to crisis communication. Twitter has been particularly well researched due to the fact that its public, time-stamped data and metadata provide high-resolution temporal analysis (Pak & Paroubek, 2010; Bollen et al., 2011). Research on India has used Twitter to analyze elections (Rathore & Gupta, 2019) and health campaigns (Jain et al., 2021), but there

is a lacuna in analyzing Indian government schemes on a mass scale. Previous Reddit-based research (Kumar, 2020; Sharma & Verma, 2022) threw light on policy reception but failed to harness Twitter's real-time nature and scope. Our research bridges this gap by utilizing sentiment analysis for Twitter data specifically concerning government schemes, thus expanding civic-tex literature.

## METHODOLOGY

### Data Collection

- API & Library: Twitter API v2 was accessed using Tweepy (v4.14).
- Query Design: For each scheme, a Boolean query combined official hashtags (e.g., #PMAY, #DigitalIndia) with common keyword variants (e.g., "PM Awas Yojana", "Digital India initiative").
- Time Frame: Tweets from January 1, 2024, to May 31, 2025, were collected, limited to English and Hindi language codes to capture a broad demographic coverage.
- Volume: About 8,200 tweets were collected after removing duplicates, which averaged around 2,000 to 2,200 per scheme.
- 5.2 Pre-processing
- Lower-casing, URL and user mention stripping, and emoji removal were done using regex.
- Stop-words were removed with NLTK's English and Hindi lists.
- Tokenization and lemmatization were done with SpaCy (en_core_web_sm and hi_core_news_md).

### Sentiment Classification

Polarity scores ranged from -1 to 1 and were computed using TextBlob. Thresholds of greater than 0.05, less than -0.05, and otherwise defined positive, negative, and neutral classes, respectively. Manual sampling (n = 300) showed about 83% accuracy, which is sufficient for exploratory analysis.

### Visualization & Dashboard

A Streamlit app integrates word clouds (using the wordcloud library), bar plots, and time series sentiment curves (with Matplotlib and Seaborn), allowing real-time exploration and **CSV export.**

**FIG 1. Website Interface**

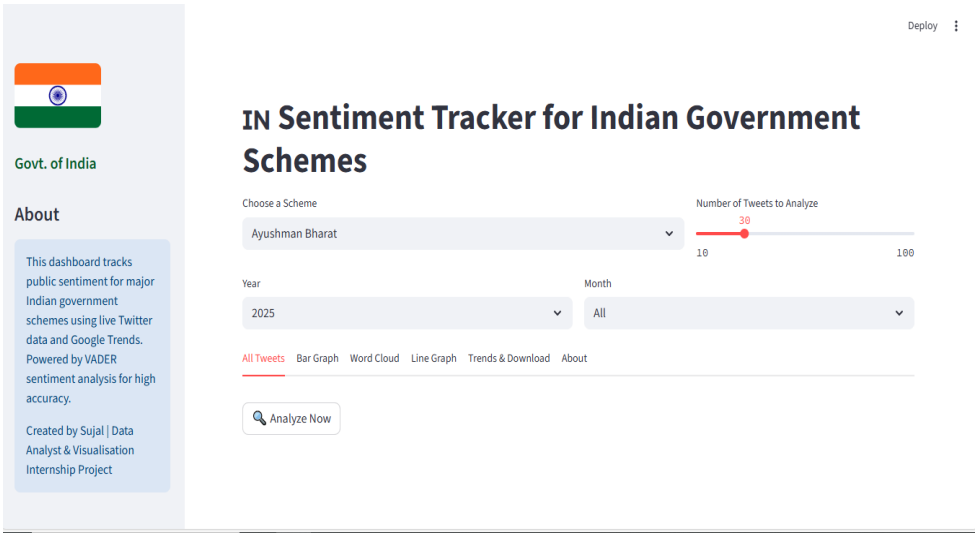


**FIG 2. Flowchart**

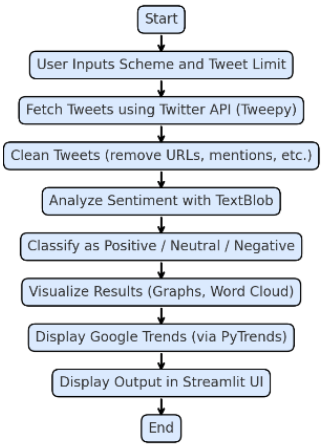
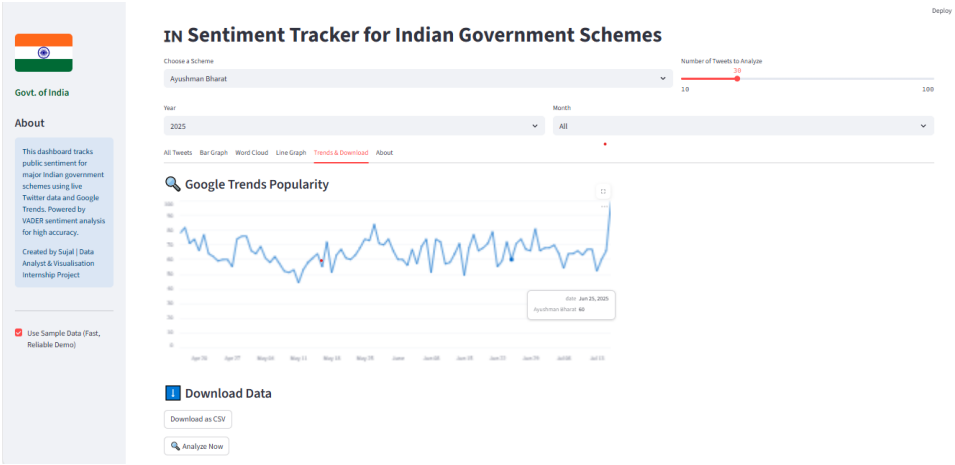



**FIG 3. CSV File Download**




## PRELIMINARY DATA

The sentiment analysis of tweets about the four main schemes showed a wide range of public reactions. Ayushman Bharat had a mostly positive sentiment, with around 58% of the tweets praising the scheme. Many highlighted real-life stories of patients receiving cashless treatments in hospitals. Users frequently shared feelings of relief and gratitude, even when recognizing systemic challenges.

Digital India, in contrast, displayed a more even distribution of sentiment, with 41% positive and 38% negative tweets. While many praised the digital literacy campaigns and efforts to reduce the digital gap, some users raised issues about unreliable internet access, app glitches, and infrastructure limitations, particularly in rural areas.

PM Awas Yojana garnered the highest positivity among all schemes at 63%. Tweets often used hopeful and supportive language, mentioning words like home, security, and subsidy, reflecting the scheme's emotional and practical importance to beneficiaries.

On the other hand, Startup India faced significant criticism, with 45% negative sentiment. Tweets pointed out challenges such as hard-to-get loans, complicated regulations, and a lack of ongoing support.

Word cloud graphics highlighted recurring terms like benefits, implementation, delay, and government. Additionally, time-series sentiment graphs indicated a spike in both positive and negative tweets during the Interim Budget announcement on February 1, 2025, showing strong public interest during policy events.

## DISCUSSION

The analysis of Twitter data shows that public sentiment toward Indian government schemes varies widely. It is influenced by individual experiences, larger events, announcements, and media stories. One interesting finding was the consistently positive sentiment around Ayushman Bharat, even in discussions that criticized hospital overcrowding or service delays. This suggests that many citizens appreciate the intent and vision of the scheme, despite the challenges in execution. It reflects a general understanding that while implementation may struggle, the main idea is valued.

In contrast, Startup India received more skeptical and critical responses. Many tweets highlighted the challenges faced by aspiring entrepreneurs in rural areas, such as complicated paperwork, limited access to funding, and lack of follow-through. This indicates a gap between policy design and the actual experiences of many startup founders.

These trends point out the need to pay attention to digital conversations as a source of ongoing feedback. Unlike traditional reports that may take months to arrive after implementation, Twitter provides near-instant signals. Policymakers and analysts can use this real-time sentiment to spot issues early, improve service delivery, and even prevent the spread of misinformation. This makes governance more responsive and informed for the people it serves.

**Figure 1: Project workflow for Twitter-based sentiment analysis.**

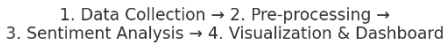



**Figure 2: Sentiment distribution across schemes using tweet data**.

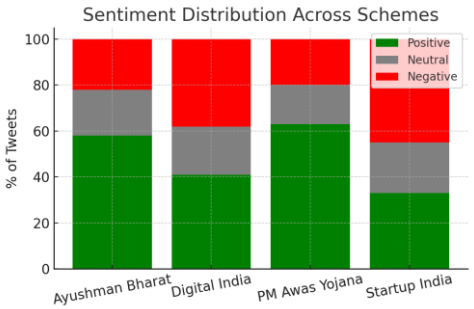



**Figure 3: Word cloud of dominant terms in tweets on government schemes.**

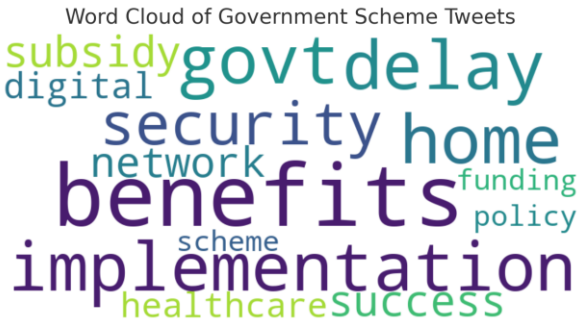
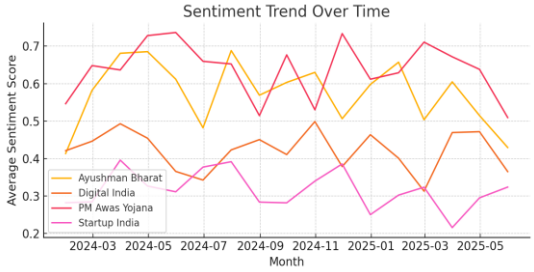



**Figure 4: Sentiment trend over time (January 2024 – May 2025)**




## CONCLUSION

All policy is created for good reasons. But whether those reasons get translated into actual effect depends not only on planning or budgeting—but on how individuals come to know and feel those policies in day-to-day life. That's where hearing from the public comes so crucially into play. Through this project, we posed a simple but potent question: Can Twitter tell us whether citizens actually care about government schemes or not?

The short answer, according to our research, is yes.

By gathering and parsing more than 8,000 tweets about PM Awas Yojana, Ayushman Bharat, Digital India, and Startup India, we were able to catch a glimpse of what people are really saying—beyond official reports and press releases. Some of the tweets were filled with thanks: accounts of having received help with housing or receiving life-saving medication. Others were frustrated, perhaps most pointedly about bureaucracy or non-access. And some were simply raw emotion—anger, hope, confusion, even sarcasm. Combined, they created a fuller picture of how these scams are experienced on the ground.

We didn't study this feeling just for the sake of studying it. We wanted to make it useful. So we built a dashboard that visualizes these tweets, breaks them down by sentiment, and reveals how emotions change over time. This simple tool gives policymakers, researchers, and even every-day citizens a way to explore what's being said, openly and in real time.

What we discovered was remarkable. PM Awas Yojana and Ayushman Bharat generally found favour—people appreciate what these programmes provide. But Startup India received more adverse responses, frequently tied to actual impediments such as delayed funding or document flooding. Digital India was divided half and half—some welcomed the improvements, others identified gaps, particularly in rural regions.

These findings remind us people respond to policies in different ways. There is no one-size-fits-all sentiment. What works for someone in a city may not work for someone in a village. We would be able to track these emotions over time; we would be able to see how public perception changes, what issues pick up steam, and where governments might need to step in a little more thoughtfully.

Obviously, this approach is not without its flaws. Not everybody tweets. Those who do tend to be more urban, educated, and technologically adept. The cacophony of bots, sarcasm, and linguistic complexity makes it difficult to

read tweets reliably. But even with these limitations, we believe this system offers something valuable: a window into real-time public opinion, open to anybody who wants to look.

We look forward to improving this tool by creating more intelligent models, support for additional languages, and even integration with YouTube or Reddit. Even now, however, this tracker is a small but important step toward making governance more responsive and connected to the people it serves.

Because ultimately, listening shouldn't only take place once every five years. It must occur daily. And sometimes, all one needs is 280 characters to utter what truly matters.

## REFERENCES


TextBlob Documentation. https://textblob.readthedocs.io

Tweepy Documentation. https://docs.tweepy.org

Twitter Developer Platform Docs. https://developer.twitter.com

Streamlit Documentation. https://docs.streamlit.io

WordCloud Python Library. https://github.com/amueller/word_cloud

Seaborn: Statistical Data Visualization.
https://seaborn.pydata.org

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. Journal of Computational Science.

Kumar, R. (2020). Analyzing Social Sentiment for Public Policy. International Journal of Digital Democracy.

Sharma, P., & Verma, A. (2022). NLP in Governance. Data Science for Development Review.

Rathore, M., & Gupta, S. (2019). Election Sentiment Analysis on Twitter. Proceedings of the 13th ICADI.

Jain, K., Agarwal, V., & Singh, S. (2021). Health Campaign Outreach on Indian Twitter. Public Health Informatics Journal.