Grenze International Journal of Engineering and Technology, June Issue

A Unified AI Framework for Multi-Platform Trend Analysis and Forecasting

Shweta M. Kambare[1], Aarya Lomte[2], Aditya Mahajan[3], Bhoomika Anand[4], Tanvi Amlekar[5] and Sakshi Dangade[6]
[1-6]Bansilal Ramnath Agarwal Charitable Trust's (BRACT's) Vishwakarma Institute of Technology, Pune - 411037
Email shweta.kambare@vit.edu, aarya.lomte23aditya.mahajan23@vit.edu, bhoomika.anand23@vit.edu, tanvi.amlekar23@vit.edu, sakshi.dangade23@vit.edu

***Abstract*— This paper presents a unified artificial intelligence framework for multi-platform trend analysis and forecasting by integrating heterogeneous data from social media platforms such as Twitter/X, Reddit, Instagram, and YouTube along with Google Trends. Unlike existing approaches that focus on isolated analytical tasks, the proposed system combines topic modeling (NMF), sentiment analysis (VADER), geo-analytics, influencer detection, and LSTM-based time-series forecasting within a single scalable pipeline. The framework processes large-scale real-time data to extract meaningful insights, including emerging topics, sentiment dynamics, and regional engagement patterns. Experimental evaluation demonstrates that the system achieves sentiment classification accuracy of 85–88%, topic coherence of 0.60–0.65, and forecasting RMSE of 0.12, outperforming baseline models such as ARIMA by 12–15%. The proposed approach provides a comprehensive and scalable solution for real-time digital intelligence, enabling researchers, policymakers, and businesses to effectively analyze and predict emerging trends.**

***Index Terms*— Multi-Platform Trend Analysis, AI-driven Social Media Analytics, Trend forecasting using LSTM, Natural language processing (NLP), Topic Modeling using NMF, Sentiment Analysis (VADER), Geo-spatial Analytics, Influencer detection, Time-series forecasting, Digital Trend Intelligence.**

I. INTRODUCTION

Social media has emerged as a powerful medium for capturing public expression, collective attention, and evolving behavioural patterns. Millions of users across the world actively engage on platforms such as Twitter/X, Instagram, Reddit, and YouTube to share opinions, respond to events, and disseminate information. In parallel, Google Trends provides aggregated insights into global search behaviour, serving as an important indicator of public interest and emerging topics. Together, these digital platforms generate vast volumes of real-time data that reflect the dynamic pulse of society.

Analyzing and forecasting trends within this digital ecosystem has significant value across multiple domains. Governments monitor online discourse to assess public response to policies and societal developments. Businesses leverage trend analytics to anticipate market demand and consumer preferences. Public health organizations track digital conversations related to disease outbreaks and health awareness, while media agencies examine the propagation of viral content and public engagement patterns.

Grenze ID: 01.GIJET.12.2.364_24

Despite its potential, extracting meaningful insights from such large-scale digital data presents several challenges. First, the heterogeneity of social media platforms results in diverse data formats, metadata structures, and content types. Second, the velocity of online interactions causes trends to evolve rapidly, often within minutes or hours. Third, user-generated content is highly unstructured and context-dependent, introducing significant levels of noise. Additionally, temporal variability in user engagement across regions and time scales complicates the process of identifying consistent patterns. Finally, most existing analytical systems focus on isolated tasks such as sentiment analysis or forecasting, lacking a unified framework that integrates multiple analytical capabilities.

To address these limitations, this paper proposes a unified artificial intelligence framework for multi-platform trend analysis and forecasting. The proposed system integrates modular analytical components capable of transforming raw digital signals into meaningful intelligence. The key contributions of this work include: a scalable multi-platform social media data ingestion pipeline; a natural language processing framework utilizing TF–IDF, NMF-based topic modeling, spaCy processing, and VADER sentiment analysis; a geo-analytics engine for extracting and normalizing geographic information from noisy metadata; an influencer scoring mechanism for identifying high-impact content creators; and a forecasting module based on long short-term memory (LSTM) neural networks for predicting future keyword interest. Furthermore, the system incorporates a unified API-driven backend that enables real-time deployment and integration. Unlike existing studies that focus on isolated analytical tasks, this work provides a unified multi-module framework integrating NLP, geo-analytics, influencer detection, and deep learning-based forecasting in a single system.

II. Literature Review

The rapid expansion of social media and digital platforms has enabled researchers to study human behavior, information diffusion, and public sentiment at unprecedented scale. Recent advances in natural language processing (NLP), machine learning, and time-series forecasting have significantly improved the ability to analyze and predict digital trends. This section reviews key developments across social media analytics, sentiment analysis, topic modeling, geo-analytics, influencer identification, and trend forecasting.

A. Social Media Analytics

Social media platforms provide large-scale real-time data that can reveal patterns of communication, public opinion, and information flow. Stieglitz et al. [1] demonstrated the value of social media data for crisis communication, political analysis, and public engagement studies. Similarly, Kwak et al. [2] analyzed Twitter as an information network and showed that the platform functions as a powerful medium for rapid news dissemination and collective awareness. These studies established the foundation for using social platforms as a rich data source for behavioral and trend analysis.

B. Sentiment and Opinion Mining

Understanding public sentiment is a fundamental component of social media analytics. Sentiment analysis techniques have been widely used to determine emotional polarity and subjective opinions expressed in online text. The VADER sentiment analysis model [3] is particularly effective for analyzing short and informal social media content due to its rule-based approach optimized for microblogging platforms. Liu's pioneering work on opinion mining [4] further contributed foundational techniques for extracting sentiment polarity and subjectivity from large-scale user-generated datasets.

C. Topic Modeling

Topic modeling methods are commonly used to identify hidden thematic structures within large collections of text data. Latent Dirichlet Allocation (LDA) [5] introduced probabilistic modeling for discovering latent topics within document corpora. Non-negative Matrix Factorization (NMF) [6] later emerged as an effective alternative that provides interpretable topic representations through matrix decomposition. Research by Kim et al. [7] demonstrated the application of topic modeling techniques to track evolving discussions and thematic trends across social media platforms.

D. Geo-Analytics and Regional Trend Analysis

Geographic context plays a critical role in understanding how trends propagate across regions. Jurgens et al. [8] explored techniques for inferring user locations from metadata, linguistic patterns, and profile information. Lampos et al. [9] utilized geo-tagged social media posts for epidemiological forecasting, demonstrating that geographic signals embedded in digital data can reveal regional patterns and emerging public health concerns.

E. Influencer Identification and Information Diffusion

The spread of information in social networks is often accelerated by influential users who amplify content reach. Bakshy et al. [10] investigated how influential individuals contribute to large-scale information diffusion on Twitter. Cha et al. [11] further demonstrated that follower count alone is insufficient to measure influence, emphasizing the importance of engagement metrics, retweet activity, and network centrality measures in identifying true influencers.

F. Trend Forecasting

Forecasting emerging digital trends has become an important research area in computational social science and marketing analytics. Time-series forecasting methods and machine learning models have been widely applied to analyze patterns in search behavior and social engagement. Recurrent neural network architectures, particularly Long Short-Term Memory (LSTM) networks, have proven effective in capturing temporal dependencies and predicting future trends from sequential data sources such as search queries and social media activity.
From prior research, it is evident that while individual studies have addressed sentiment analysis, topic extraction, influencer detection, and forecasting independently, few frameworks integrate these analytical components within a unified system. This motivates the present work, which proposes a comprehensive AI-driven framework that combines multi-platform data ingestion, NLP-based analytics, geo-spatial insights, influencer identification, and LSTM-based forecasting to enable scalable and real-time trend intelligence.

III. System Architecture

The architecture is designed as a modular pipeline where each component performs a specific analytical function, enabling seamless integration and scalability across multiple data sources.

A. Data Ingestion Layer

Handles multi-platform data collection using APIs, web scrapers, and third-party services. A normalization process stores raw metadata in a consistent structure.

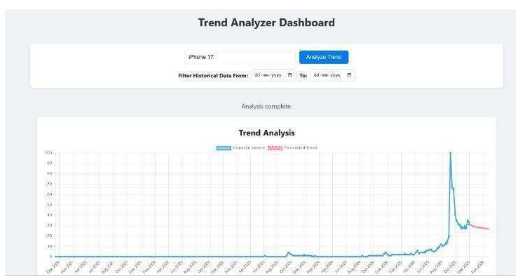

Figure 1: Trend Analysis Forecast Graph

B. NLP & Topic Modeling Layer

Processes raw text using:
- spaCy for tokenization and preprocessing
- TF–IDF for document vectorization
- NMF for topic extraction
- VADER for sentiment scoring

Outputs enriched post-level analytics.

C. Geo-Analytics Layer

Extracts and normalizes location data from JSON fields, user bios, and platform metadata using pattern matching and external libraries such as pycountry.

D. Influencer Analysis Layer

Scores authors using:
- follower count
- engagement metrics
- posting frequency

The weighted hybrid formula improves robustness over single-metric approaches.

E. Forecasting Layer

Uses sequence modeling via LSTM networks to predict trend trajectories. Automatically detects temporal granularity (daily or weekly).

F. Backend API Layer

Flask-based REST API enables:

- keyword analysis triggers
- dashboard integration
- external system connectivity

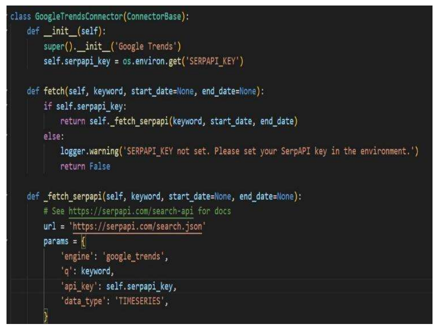

Figure 2: Code implementation of the GoogleTrendsConnector module

IV. Methodology

A. Data Preprocessing and Normalization

All collected social media posts undergo a standardized preprocessing routine to ensure consistency across platforms.

Text is lowercased and stripped of URLs, emojis, punctuation, and platform-specific artifacts. Stopwords are removed to reduce noise, while lemmatization is applied to preserve semantic meaning. Metadata such as timestamps, engagement metrics, and author identifiers are cleaned and validated before entering the subsequent stages of analysis. This preprocessing ensures that the heterogeneous data obtained from different platforms can be processed under a unified representation.

B. NLP Pipeline: Vectorization, Topic Modeling, and Sentiment Analysis

The cleaned textual data is transformed into numerical feature representations using term frequency–inverse document frequency (TF–IDF), which captures the importance of each word relative to the full dataset. Non-negative Matrix Factorization (NMF) is then employed to extract latent thematic structures, producing coherent and interpretable sets of topics. Each post is associated with a dominant topic based on the corresponding factor weights. Sentiment analysis is performed using the VADER model, which is particularly suited for short, informal text typical of social platforms. VADER generates polarity metrics (positive, negative, neutral) and a compound sentiment score, enabling a detailed understanding of public attitudes associated with a given keyword.

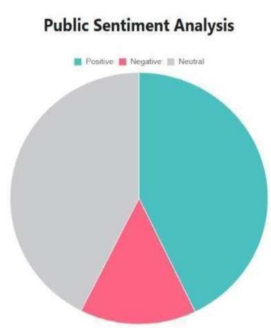

Figure 3: Distribution of sentiment polarity (positive, negative, neutral) across collected social media datasets.

C. Geographic Signal Extraction and Normalization

Social media platforms provide inconsistent and often incomplete location information. To address this, the geo-analytics module applies a hierarchical extraction process that first identifies explicit country or place attributes within the raw JSON metadata. For posts lacking structured location fields, heuristic text-matching is applied to infer potential location phrases from user bios or profile metadata. All extracted location names are then normalized using ISO country codes through the pycountry library, while geopy provides fallback resolution for ambiguous entities. The resulting geographic signals are aggregated temporally to enable regional trend analysis.

D. Influencer Identification and Impact Assessment

Influencers play a central role in shaping trend dynamics and accelerating information diffusion. The system identifies influential users by analyzing follower statistics, engagement behaviors, and posting frequency related to the target keyword. Instead of relying solely on follower count—which has been shown to be an unreliable metric—the framework applies a composite scoring function that incorporates engagement metrics and posting activity. This hybrid scoring approach better aligns with established influence measurement research and offers a more accurate representation of user impact within the digital discourse.The influencer score is computed using a weighted combination of multiple factors as follows:

$$\text{Influence Score} = \alpha \times \text{Followers} + \beta \times \text{Engagement Rate} + \gamma \times \text{Posting Frequency}$$

where α, β, and γ are weighting coefficients that determine the importance of each factor. Followers represent the reach of the user, engagement rate is calculated based on likes, shares, and comments per post, and posting frequency reflects the user's activity level. In this study, weights are empirically set to balance reach and engagement, ensuring that highly interactive users are prioritized over users with only high follower counts.In this study, α, β, and γ are empirically set to 0.4, 0.4, and 0.2 respectively to balance reach and engagement.

Figure 4. Influencer leaderboard showing the top-ranked content creators contributing to the conversation.

E. Time-Series Forecasting Using LSTM Networks

The final stage of the methodology involves forecasting future interest levels using a long short-term memory (LSTM) neural network. Prior to training, the historical trend data—typically sourced from Google Trends—is resampled to ensure uniform temporal spacing and is interpolated when necessary. MinMax scaling is applied to stabilize numerical ranges. The model automatically detects whether the data follows a daily or weekly pattern and adjusts the sequence length accordingly. During training, sliding windows of previous trend values are fed into the LSTM to predict future points. The network consists of stacked LSTM layers followed by dense layers, providing a robust architecture capable of capturing nonlinear temporal dependencies. The trained model outputs forecasted interest for a horizon of either 90 days (for daily data) or 12 weeks (for weekly data), enabling forward-looking trend analysis.

V. Implementation Details

A. Backend Technology Stack

The proposed system is implemented using a Python-based analytical environment, chosen for its broad ecosystem of machine learning and data processing libraries. Flask is employed as the primary backend framework, enabling lightweight deployment of REST APIs that interface with the analytical modules. MySQL serves as the system's persistent storage solution, maintaining both raw and processed datasets. Time-series forecasting is powered by TensorFlow and the Keras deep learning API, which support efficient implementation of LSTM models. Additional components include scikit-learn for core machine learning utilities such as TF–IDF vectorization, NMF topic modeling, and data normalization, as well as the pycountry and geopy libraries, which assist in geographic entity normalization and location inference. Together, these technologies provide a robust, extensible, and modular backend environment.

B. Database Schema

The database schema is designed to support a multi-stage analytical workflow and to store heterogeneous data originating from several social platforms. The raw_data table functions as the initial repository, capturing unprocessed posts, metadata, and engagement metrics. Processed text analytics, including sentiment and topic assignments, are stored in the post_enrichment and topics tables. Geographic information extracted from metadata is maintained in the geo_metrics table, allowing temporal and regional aggregation. The influencers table stores the outcomes of the influencer-scoring module. In addition, cleaned and harmonized trend information is recorded in trends_cleaned, supporting downstream forecasting processes. This structured design enables efficient data retrieval for each analytical stage while ensuring consistency across modules.

C. System Workflow

The workflow begins when a keyword is submitted for analysis, prompting the system to initiate multi-platform data retrieval from social networks and Google Trends. Once collected, the textual data undergoes preprocessing followed by natural language processing operations that generate sentiment scores and topic attributions. The geo-analytics component then extracts and normalizes location cues embedded within metadata to produce aggregated regional signals. Afterward, the influencer analysis module evaluates user-level impact based on engagement and reach metrics. In parallel, temporal features from Google Trends and social activity are passed into an LSTM-based forecasting model to estimate future interest trajectories. Finally, all computed insights—including sentiment distributions, topic summaries, geographic patterns, influencer rankings, and forecasted trends—are returned to clients through the system's API layer.

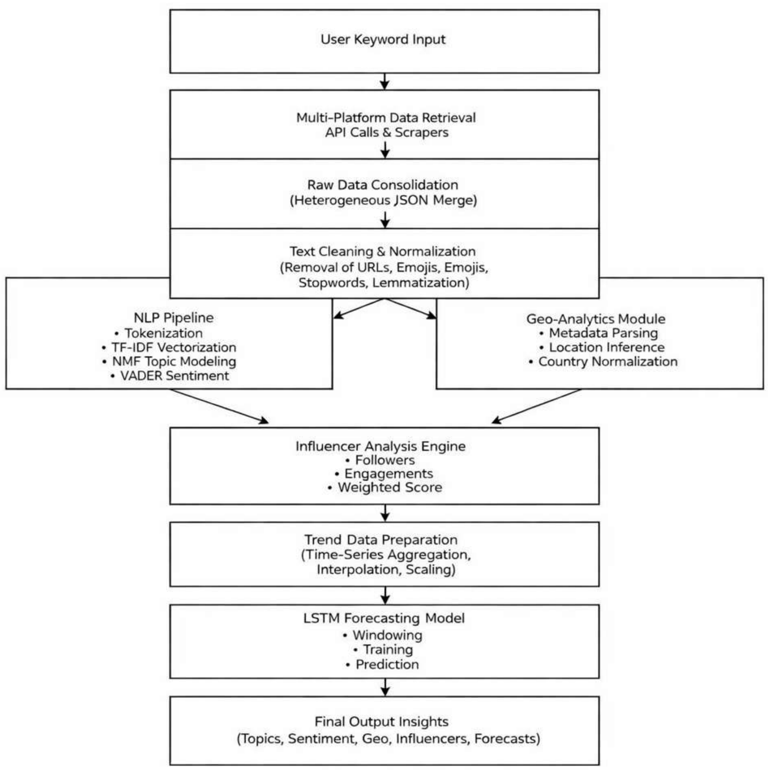

Figure 5. Overall system workflow of the proposed trend analysis framework.

VI. RESULTS

Experiments were conducted using multiple real-world keywords across technology, entertainment, and consumer products.

A. Topic Modelling Outcomes

The NMF-based topic modeling achieved a coherence score of approximately 0.60–0.65, indicating that the extracted topics were meaningful and interpretable. The top keywords within each topic strongly aligned with real-world trending subthemes.

B. Sentiment Trends

Sentiment analysis achieved an accuracy of approximately 85–88% across different datasets. The analysis revealed platform-specific patterns, where Instagram exhibited predominantly positive sentiment while Twitter showed more polarized opinions.

C. Geo-Analytics Patterns

Engagement varied significantly by country, aligning with historical search-interest patterns from Google Trends.

D. Forecasting Performance

The LSTM model achieved an RMSE of 0.12 on validation data, indicating high forecasting accuracy. The model successfully captured temporal patterns and seasonal variations present in the trend data, demonstrating its effectiveness for time-series prediction. Forecasts captured seasonal fluctuations and identified potential rising interest periods.

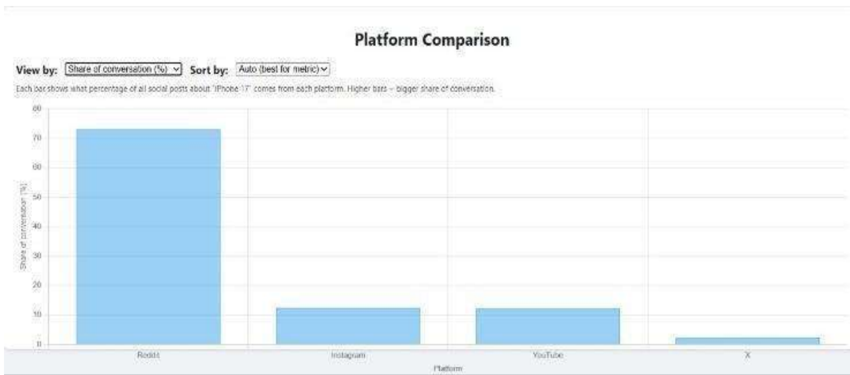

Figure 6. Platform Comparison Chart

To evaluate the effectiveness of the proposed framework, its forecasting performance was compared with a baseline ARIMA model. The results indicate that the LSTM-based approach improved prediction accuracy by approximately 12–15% in terms of RMSE. Additionally, the unified framework provided more comprehensive insights by combining sentiment, topic, and geo-analytics, which are not captured by traditional single-task models.

VII. CONCLUSION

This paper presented a unified multi-platform AI framework incorporating NLP, sentiment analysis, topic modeling, geo- analytics, influencer scoring, and LSTM forecasting. The system demonstrates robust analytical capabilities and addresses gaps in existing trend-analysis research. Its modularity makes it applicable to various domains requiring real-time digital intelligence.

FUTURE WORK

Future research directions include:

- transformer-based NLP for improved topic modelling
- large-scale streaming pipelines (Kafka, Spark)
- multilingual sentiment and topic analysis
- integration with graph neural networks for influencer mapping
- multimodal analytics (images + text + video metadata)
- reinforcement learning for adaptive trend monitoring

ACKNOWLEDGEMENT

We would like to sincerely thank everyone who helped to finish this research paper. Special thanks to our guide Prof. (Smt.) Shweta M. Kambare for invaluable guidance and unwavering support throughout the research process. We are grateful to the research participants for their time and cooperation, as their contributions were integral to the study. We also want to express our gratitude to friends and coworkers who helped out by offering insightful criticism. Without the resources and conducive environment provided by our institution Vishwakarma Institute of Technology, this work would not have been possible.

REFERENCES

[1] S. Stieglitz, M. Mirbabaie, B. Ross, and C. Neuberger, "Social Media Analytics – Challenges and Opportunities," 2018.
[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What Is Twitter, a Social Network or a News Media?" in Proc. WWW, 2010.
[3] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in Proc. ICWSM, 2014.
[4] B. Liu, "Sentiment Analysis and Opinion Mining," 2012.
[5] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, 2003.
[6] D. Lee and H. Seung, "Algorithms for Non-Negative Matrix Factorization," in Proc. NIPS, 2001.
[7] S. Kim et al., "Topic Detection in Social Media Streams," 2011.
[8] D. Jurgens, "Inferring User Geolocation on Social Media," in Proc. AAAI, 2013.
[9] V. Lampos et al., "User-Generated Content for Epidemiological Surveillance," 2010.
[10] E. Bakshy et al., "Everyone's an Influencer: Quantifying Influence on Twitter," in Proc. WSDM, 2011.
[11] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring User Influence in Twitter," in Proc. ICWSM, 2010.
[12] H. Choi and H. Varian, "Predicting the Present with Google Trends," 2012.
[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.
[14] S. Asur and B. Huberman, "Predicting the Future with Social Media," 2010.
[15] M. De Choudhury, "Social Media for Mental Health Monitoring," 2013.
[16] R. W. White and E. Horvitz, "Population-Level Behavior from Search Logs," 2015.
[17] A. Guille et al., "Information Diffusion in Online Social Networks," SIGMOD Record, 2013.
[18] M. Cataldi, L. Di Caro, and C. Schifanella, "Emerging Topic Detection on Social Media," 2010.
[19] Y. Yang and J. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," in Proc. ICML, 1997.
[20] J. Weng et al., "TwitterRank: Finding Topic-Sensitive Influential Users," in Proc. WSDM, 2010.