International Journal of Scientific Research in Engineering & Technology
Volume6, Issue3 (May-June 2026), PP: 222-229.
https://www.doi.org/10.59256/ijsreat.20260603033
www.ijsreat.com
ISSN No: 2583-1240

Early Detection and Forecasting of Emerging Trends Using Explainable Machine Learning and Multi-Source Social Media Analytics

Bhagyashree Nishane[1], Charuta Khadke[2], Dr. Deepali Y. Kirange[3]
[1,2,3] Assistant Professor, KCES's Institute of Management and Research, Jalgaon, Maharashtra, India.

To Cite this Article: *Bhagyashree Nishane[1], Charuta Khadke[2], Dr. Deepali Y. Kirange[3], "Early Detection and Forecasting of Emerging Trends Using Explainable Machine Learning and Multi-Source Social Media Analytics", International Journal of Scientific Research in Engineering & Technology, Volume 06, Issue 03, May-June 2026, PP:222-229.*

Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: *The accelerating pace at which topics rise, peak, and fade across digital platforms has made early trend detection a critical capability for journalism, public health surveillance, marketing, and policy planning. Conventional trend-detection systems rely on single-platform signals and opaque deep learning models, which limits both their coverage and their trustworthiness in operational decision-making contexts. This paper proposes a multi-source social media analytics framework that fuses textual, temporal, and network signals from Twitter/X, Reddit, YouTube comments, and news RSS feeds to detect emerging trends substantially earlier than single-source baselines. The framework combines a burst-aware temporal feature extractor, a graph-based diffusion model capturing cross-platform propagation, and a gradient-boosted forecasting ensemble, with SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) layers providing per-prediction interpretability. We evaluate the framework on a curated 14-month, multi-platform dataset spanning approximately 9.2 million posts across 1,800 manually and weakly labeled emerging topics. The proposed system achieves a mean detection lead time of 36.4 hours ahead of trend-platform native trending lists, an F1-score of 0.91 for trend/non-trend classification, and a mean absolute percentage error (MAPE) of 11.3% for 48-hour volume forecasting, outperforming LSTM, Prophet, and single-platform XGBoost baselines. Explainability analysis shows that burst acceleration, cross-platform co-occurrence, and influencer-adjusted reach are the most consistently influential features across topic categories. The results demonstrate that combining multi-source fusion with explainable machine learning yields both earlier and more trustworthy trend forecasts, with direct implications for misinformation early-warning systems, public health surveillance, and real-time marketing intelligence.*

Key Words: *explainable AI, SHAP, LIME, social media analytics, multi-source data fusion, time-series forecasting.*

I. INTRODUCTION

Social media platforms have become the earliest and most sensitive sensors of collective attention, often registering the first signals of emerging news events, public health concerns, product issues, and cultural phenomena hours or days before they are picked up by mainstream media or appear in conventional analytics dashboards. The ability to detect such trends early, and to forecast their likely trajectory, has direct value for crisis response teams, brand managers, journalists, and public health authorities. However, the same properties that make social media data valuable for early detection, namely its volume, velocity, and heterogeneity, also make it difficult to model reliably.

Most existing trend-detection systems suffer from three limitations. First, they are typically single-platform, relying exclusively on Twitter/X or a single forum, which leaves them blind to trends that originate or accelerate on other platforms before crossing over. Second, the majority of recent high-performing systems are built on deep neural architectures whose internal decision logic is opaque, making it difficult for analysts to understand why a particular topic was flagged as an emerging trend, which in turn limits trust and adoption in operational settings. Third, many systems conflate detection (identifying that something unusual is happening) with forecasting (predicting how the topic will evolve), even though the two tasks require different modeling assumptions and evaluation criteria.

This paper addresses these gaps by proposing an integrated framework that (i) fuses signals from multiple social media and news sources into a unified feature space, (ii) explicitly separates early detection from short-horizon forecasting while sharing a common feature backbone, and (iii) embeds explainability directly into both stages using SHAP and LIME, so that every flagged trend is accompanied by a human-interpretable rationale. The framework is designed to be platform-agnostic at the ingestion layer, allowing new data sources to be added without retraining the core detection model.

1.1 Motivation and Problem Statement

Three motivating scenarios illustrate the need for this work. In public health, early identification of symptom-related chatter clusters can provide lead time ahead of formal case reporting. In misinformation monitoring, the cross-platform diffusion pattern of a narrative is often a stronger signal of coordinated inauthentic behavior than its content alone. In commercial marketing

intelligence, identifying a product or brand trend while it is still in its acceleration phase, rather than after it peaks, materially changes the value of a response. In all three scenarios, a black-box alert is less actionable than one accompanied by an explanation of which signals drove the detection.

1.2 Research Objectives

- To design a multi-source data fusion pipeline that harmonizes heterogeneous social media and news signals into a common temporal-textual-network feature representation.
- To develop a burst-aware early detection model that identifies emerging trends ahead of platform-native trending mechanisms.
- To build a short-horizon forecasting ensemble that predicts topic volume and sentiment trajectory for 24- to 72-hour windows.
- To integrate SHAP- and LIME-based explainability layers that provide consistent, human-readable justifications for both detection and forecasting outputs.
- To empirically evaluate the framework against established single-source and black-box baselines on lead time, classification accuracy, forecasting error, and explanation fidelity.

1.3 Contributions

- A cross-platform feature fusion architecture combining lexical, temporal-burst, and network-diffusion signals from four heterogeneous source types.
- A two-stage detection-forecasting pipeline that decouples the binary emergence decision from continuous trajectory prediction while sharing a common encoder.
- An explainability layer that combines global SHAP summaries with local LIME explanations and a novel explanation-stability metric for evaluating interpretability under streaming conditions.
- A large-scale empirical evaluation on a 14-month, 9.2-million-post multi-platform corpus with comparisons against six baseline methods.

1.4 Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work in trend detection, social media analytics, and explainable machine learning. Section 3 describes the proposed system architecture and data fusion methodology. Section 4 details the explainability components. Section 5 describes the experimental setup, dataset, and evaluation metrics. Section 6 presents results and discussion. Section 7 discusses limitations and threats to validity. Section 8 concludes the paper and outlines directions for future work.

II. LITERATURE REVIEW

2.1 Trend Detection on Social Media

Early work on social media trend detection focused primarily on burst detection in single-platform keyword streams, using statistical change-point methods to flag sudden increases in term frequency. Subsequent research extended these approaches with topic-modeling techniques such as Latent Dirichlet Allocation and its dynamic variants, which allowed trend detection systems to track evolving themes rather than fixed keywords. More recent work has incorporated graph-based diffusion models that treat trend propagation as an epidemic-like spreading process across a social network, borrowing techniques from computational epidemiology to estimate reproduction-number-like quantities for information spread.

2.2 Multi-Source and Cross-Platform Analytics

A growing body of literature has recognized that single-platform analysis systematically underestimates the breadth and speed of trend emergence, since many topics originate on smaller or niche platforms before crossing over to mainstream networks. Cross-platform studies have examined how narratives migrate between forums, microblogging sites, and video-comment sections, generally finding that diffusion velocity and cross-platform co-occurrence are stronger predictors of eventual virality than within-platform engagement metrics alone. However, most of these studies remain descriptive or retrospective, with comparatively few operational early-warning systems built on multi-source fusion.

2.3 Time-Series Forecasting for Social Signals

Forecasting the volume and sentiment trajectory of social media activity has been approached using classical statistical models, exponential smoothing variants, and more recently sequence-based deep learning architectures including LSTM, GRU, and Transformer-based encoders. While deep sequence models generally achieve lower forecasting error on aggregate benchmarks, several studies have noted that their gains diminish substantially in early-stage, low-volume regimes, which are precisely the conditions under which early detection is most valuable and most difficult.

2.4 Explainable Machine Learning

Explainable AI methods, particularly SHAP and LIME, have been widely adopted in tabular and structured-prediction settings to provide feature-level attributions for individual predictions. SHAP's foundations in cooperative game theory give it desirable consistency properties, while LIME's local surrogate-model approach offers computational efficiency at some cost to global consistency. Applications of explainability to streaming or temporal data remain comparatively underexplored, with most existing work treating explanation generation as a static, post-hoc step rather than as a continuously updated component of a live monitoring system.

2.5 Explainability in Social Media Analytics

A smaller set of studies has begun to apply explainability techniques specifically to social media classification tasks such as rumor detection, hate-speech identification, and bot detection, generally reporting that practitioners trust and act on system outputs more readily when attributions are provided. However, very few studies have extended explainability to the trend-forecasting task itself, and fewer still have evaluated whether explanations remain stable as the underlying data stream evolves, which is the gap this paper targets directly.

2.6 Research Gap

Across the reviewed literature, three gaps recur consistently: (1) trend detection systems are predominantly single-source, limiting their effective lead time; (2) forecasting models that achieve strong accuracy are rarely paired with interpretability mechanisms suitable for operational analysts; and (3) explainability techniques, where present, are typically applied retrospectively rather than as an integrated, streaming-compatible component. The framework proposed in this paper is designed to close all three gaps simultaneously within a single architecture.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed framework consists of four layers: (1) a multi-source ingestion and normalization layer, (2) a feature fusion layer combining lexical, temporal-burst, and network-diffusion signals, (3) a two-stage detection-forecasting model, and (4) an explainability layer that operates on top of both stages. Figure 1 (described conceptually below, as this is a text-based methodology section) summarizes the end-to-end pipeline from raw multi-platform streams to explained trend alerts.

3.1 Data Sources and Ingestion

Four source types are ingested continuously: (i) microblogging posts from Twitter/X via the filtered stream API, (ii) Reddit submissions and comments from a curated set of subreddits spanning news, technology, health, and entertainment, (iii) YouTube video metadata and top-level comments for videos matching tracked seed keywords, and (iv) news RSS feeds from approximately 120 outlets. Each source is normalized into a common schema containing timestamp, platform identifier, author pseudonymized identifier, raw text, engagement counts, and, where available, geolocation at the country level.

3.2 Topic Identification and Clustering

Incoming text is embedded using a sentence-transformer encoder and clustered using an online variant of HDBSCAN that supports incremental cluster assignment as new documents arrive. Each resulting cluster is treated as a candidate topic. Candidate topics are merged across platforms when their centroid embeddings exceed a cosine similarity threshold of 0.78 and their temporal windows overlap, enabling cross-platform topic linking without requiring identical keywords.

3.3 Feature Fusion Layer

For each candidate topic at each time step, three feature groups are computed and concatenated into a unified feature vector.

3.3.1 Lexical and Semantic Features

- TF-IDF weighted n-gram vectors reduced via truncated SVD to 64 dimensions.
- Mean sentence-transformer embedding of the topic cluster's constituent documents.
- Sentiment polarity and subjectivity scores aggregated at the topic level.

3.3.2 Temporal-Burst Features

- Kleinberg burst-detection state and burst intensity score.
- First and second derivatives of post-volume time series (velocity and acceleration).
- Inter-arrival time entropy, capturing whether posting activity is bursty or steady.

3.3.3 Network-Diffusion Features

- Cross-platform co-occurrence count within a rolling 6-hour window.
- Influencer-adjusted reach, weighting each post by the log-scaled follower count of its author.
- Estimated diffusion reproduction number, computed from a discrete-time branching-process approximation of retweet/share cascades.

3.4 Two-Stage Detection-Forecasting Model

The detection stage is framed as a binary classification problem: given the feature vector at time t, predict whether the topic will exceed a platform-relative emergence threshold within the following 24 hours. A gradient-boosted decision tree ensemble (LightGBM) is used for this stage because tree ensembles retain strong performance on tabular, heterogeneous feature sets while remaining naturally compatible with SHAP's TreeExplainer for exact, efficient attribution computation.

The forecasting stage takes topics flagged as emerging and predicts the post-volume and sentiment trajectory over the subsequent 24, 48, and 72-hour horizons. This stage uses a hybrid ensemble combining an XGBoost regressor on the fused tabular features with a lightweight temporal convolutional network operating directly on the raw volume time series, with final predictions produced by a weighted average whose weights are tuned on a held-out validation window.

3.5 Training and Validation Strategy

All models are trained using rolling-origin (walk-forward) validation to respect the temporal ordering of the data and to avoid information leakage from future trend behavior into earlier training folds. Class imbalance in the detection stage, where genuinely emerging topics are a small minority of all candidate clusters, is addressed using a combination of class-weighted loss and synthetic minority oversampling applied only within each training fold.

IV. EXPLAINABILITY LAYER

Interpretability is treated as a first-class architectural component rather than a post-hoc add-on. Two complementary techniques are integrated: SHAP for globally consistent, theoretically grounded attributions, and LIME for fast, locally faithful explanations suitable for streaming throughput constraints.

4.1 SHAP-Based Global and Local Attribution

Because both the detection and forecasting stages rely on tree-based ensembles, SHAP's TreeExplainer is used to compute exact Shapley values in polynomial time rather than relying on sampling-based approximations. Global feature importance is obtained by averaging absolute SHAP values across all topics in a rolling 7-day window, while local, per-alert explanations are generated at the moment each trend alert is emitted, listing the top contributing features and their directional effect on the emergence probability.

4.2 LIME-Based Local Surrogate Explanations

LIME is used as a complementary, model-agnostic cross-check, particularly for the temporal convolutional network component of the forecasting ensemble, where exact SHAP computation is more expensive. A local linear surrogate model is fit around each forecast using perturbed neighbors in feature space, yielding an explanation that is compared against the SHAP-based attribution for consistency.

4.3 Explanation Stability Metric

To assess whether explanations remain trustworthy as the underlying data stream evolves, this paper introduces an explanation-stability metric, defined as the rank correlation between the top-k feature attributions for the same topic cluster computed at consecutive time steps. High stability indicates that the model's reasoning is not erratically shifting as new data arrives, which is essential for analyst trust in a live monitoring setting.

4.4 Human-Readable Alert Generation

Each emitted alert is automatically rendered into a short natural-language justification template populated from the top SHAP features, for example summarizing that a topic was flagged primarily due to a sharp increase in posting velocity combined with simultaneous appearance across two or more platforms within a six-hour window. This translation step is intended to make the explanation layer usable by non-technical analysts.

V.EXPERIMETAL SETUP

5.1 Dataset Description

The evaluation dataset spans fourteen months and combines four source types as summarized in Table 1. A total of 1,800 candidate topic clusters were labeled as emerging or non-emerging using a combination of manual annotation by three independent raters (with disagreements resolved by majority vote) and weak labeling derived from subsequent appearance on platform-native trending lists or mainstream news coverage within 72 hours.

Source	Documents	Date Range	Avg. Posts/Day
Twitter/X	4,120,000	Mar 2025–Apr 2026	9,780
Reddit	2,860,000	Mar 2025–Apr 2026	6,790
YouTube Comments	1,540,000	Mar 2025–Apr 2026	3,650
News RSS	680,000	Mar 2025–Apr 2026	1,610
Total	9,200,000	14 months	21,830

Table 1. Summary statistics of the multi-source dataset used for evaluation

5.2 Baseline Methods

- Single-platform XGBoost: the proposed tabular features computed from Twitter/X data only.
- LSTM: a sequence model operating on raw per-platform volume time series without fused features.
- Prophet: an additive time-series decomposition model applied to aggregate volume.
- BERT-based classifier: a fine-tuned transformer text classifier without explicit temporal or network features.
- Graph Neural Network (GNN) diffusion model: network-only features without lexical or burst signals.
- Proposed framework (ablations): full model, and variants removing each feature group individually to assess contribution.

5.3 Evaluation Metrics

- Detection lead time: the time difference between the proposed system's alert and the corresponding platform-native trending list appearance.
- Precision, Recall, and F1-score for the binary emergence classification task.
- Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) for 24/48/72-hour volume forecasts.
- Explanation fidelity: agreement between SHAP and LIME top-3 feature rankings, measured via Spearman rank correlation.
- Explanation stability: as defined in Section 4.3, computed across consecutive 1-hour time steps.

5.4 Implementation Details

All models were implemented in Python 3.11 using LightGBM 4.x, XGBoost 2.x, PyTorch 2.x for the temporal convolutional network, and the official SHAP and LIME libraries. Hyperparameters for all gradient-boosted models were tuned via Bayesian optimization over a held-out validation window comprising the final two months of the dataset, with the preceding twelve months used for rolling-origin training.

VI.RESULTS

6.1 Detection Performance

Table 2 reports precision, recall, F1-score, and mean detection lead time for the proposed framework against all baselines. The full framework achieves the highest F1-score (0.91) and the longest mean lead time (36.4 hours), outperforming the strongest single-feature-group baseline, the GNN diffusion model, by 10 percentage points in F1-score and nearly 15 hours in lead time. The single-platform XGBoost baseline, despite using the same modeling family as the proposed detection stage, trails substantially, confirming that the performance gain is attributable primarily to multi-source fusion rather than model choice alone.

Model	Precision	Recall	F1-score	Lead (h)
Single-platform XGBoost	0.79	0.74	0.76	14.2
LSTM (volume only)	0.75	0.71	0.73	11.8
Prophet	0.68	0.64	0.66	8.4
BERT classifier (text only)	0.81	0.77	0.79	13.5
GNN diffusion model	0.80	0.83	0.81	21.6
Proposed (full framework)	0.92	0.90	0.91	36.4

Table 2. Detection performance and mean lead time across models.

6.2 Ablation Analysis

Table 3 presents an ablation study removing each feature group from the full framework individually. Removing temporal-burst features produces the largest performance drop (F1-score falling from 0.91 to 0.83, lead time falling from 36.4 to 22.9 hours), indicating that burst dynamics are the single most important feature group for early detection. Removing network-diffusion features produces the second-largest drop, while removing lexical and semantic features has a comparatively smaller, though still measurable, effect.

Configuration	F1-score	Lead Time (h)
Full framework	0.91	36.4
− Network-diffusion features	0.85	27.1
− Temporal-burst features	0.83	22.9
− Lexical/semantic features	0.86	33.0
Single-source (Twitter/X only)	0.79	14.2

Table 3. Ablation study isolating the contribution of each feature group.

6.3 Forecasting Accuracy

Table 4 reports MAPE for 24-, 48-, and 72-hour volume forecasts. The proposed hybrid ensemble achieves the lowest error at all three horizons, with the performance gap widening at longer horizons, suggesting that the fused multi-source representation degrades more gracefully than single-source or purely sequence-based alternatives as the forecasting window extends.

Model	24h MAPE	48h MAPE	72h MAPE
Prophet	19.8%	24.6%	31.2%
LSTM	15.4%	19.7%	25.8%
Single-platform XGBoost	13.9%	17.2%	22.5%
Proposed (full framework)	8.1%	11.3%	16.7%

Table 4. Forecasting error (MAPE) across horizons.

6.4 Explainability Results

Across all evaluated topic categories, burst acceleration, cross-platform co-occurrence, and influencer-adjusted reach consistently ranked among the top five SHAP-attributed features, together accounting for an average of 58% of total absolute SHAP value. SHAP and LIME top-3 feature rankings showed a mean Spearman rank correlation of 0.74, indicating reasonably strong but imperfect agreement between the two explanation methods, with disagreement concentrated in topics exhibiting unusually sparse early-stage data. The explanation-stability metric averaged 0.81 across the evaluation window, with the lowest stability observed during the first two hours after a topic's first appearance, when feature estimates are most sensitive to incoming data.

6.5 Category-Level Analysis

Disaggregating results by topic category (health, technology, entertainment, and politics) showed that network-diffusion features were disproportionately important for politically themed topics, consistent with the coordinated sharing patterns often observed in that domain, while lexical-semantic features carried relatively more weight for entertainment topics, where novel terminology and named entities emerge rapidly.

VII.DISCUSSION

7.1 Interpretation of Findings

The results support the central hypothesis of this paper: combining multi-source fusion with an explicit detection-forecasting separation yields both earlier and more accurate trend identification than single-source or undifferentiated approaches. The ablation results further suggest that temporal-burst dynamics, rather than raw textual content, are the primary driver of early lead time, which has a practical implication for system design: investment in robust, low-latency burst-feature computation may yield greater returns than investment in increasingly large language-model-based text encoders.

7.2 Practical Implications

The explainability results indicate that the framework's alerts are not only earlier but also accompanied by attributions that are reasonably stable and consistent across two independent explanation methods, which is a precondition for analyst trust in operational deployment. The observed reduction in explanation stability during the first two hours after topic emergence suggests that operational systems should treat very early alerts with appropriately calibrated confidence, perhaps via a stability-aware confidence display, rather than presenting all alerts with uniform certainty.

7.3 Comparison with Prior Work

The lead-time improvement reported here is broadly consistent with the directional findings of prior cross-platform diffusion studies, which have generally observed that cross-platform signals precede single-platform virality. This paper extends that line of work by quantifying the lead-time benefit in an end-to-end detection-forecasting system rather than in a purely descriptive analysis, and by pairing the performance gain with a systematic interpretability evaluation, which prior cross-platform studies have largely not addressed.

VIII.LIMITATIONS AND THREAD TO VALIDITY

- Platform coverage: the framework currently ingests four source types; emerging trends originating on platforms outside this set (e.g., messaging-app forwards) are not observable and would not be detected.
- Labeling bias: weak labels derived from subsequent appearance on platform-native trending lists may themselves reflect platform-specific algorithmic curation rather than an objective ground truth of emergence.
- Temporal generalization: the fourteen-month evaluation window, while substantial, may not capture longer-cycle seasonal effects or platform policy changes that could alter diffusion dynamics in ways not represented in the training data.
- Explanation faithfulness: SHAP and LIME, while widely used, are approximations to true model reasoning, and the moderate agreement observed between them (Spearman correlation 0.74) indicates that some caution is warranted in treating either method's attributions as a complete account of model behavior.
- Computational cost: the full multi-source pipeline, including exact SHAP computation, requires more compute per topic than single-source baselines, which may constrain deployment in extremely high-throughput settings without further optimization.

IX.CONCLUSION AND FUTURE WORK

This paper presented a multi-source social media analytics framework for early trend detection and short-horizon forecasting, with explainability integrated as a core architectural component rather than a post-hoc addition. By fusing lexical, temporal-burst, and network-diffusion signals across four heterogeneous source types, the proposed framework achieved a mean detection lead time of 36.4 hours and an F1-score of 0.91, substantially outperforming single-source and black-box baselines, while SHAP- and LIME-based explanations provided consistent and reasonably stable justifications for individual alerts.

Future work will extend the framework along three directions: incorporating additional source types such as messaging-platform public channels and image-based content via multimodal embeddings; developing adaptive, uncertainty-aware alert thresholds that adjust based on the explanation-stability metric introduced in this paper; and conducting a field deployment study to assess analyst trust and decision quality when using explainable trend alerts in a live operational setting rather than in retrospective evaluation.

9.1 Ethical Considerations

Because early trend detection systems can be applied to surveillance-adjacent use cases, deployments of this framework should incorporate clear data minimization practices, pseudonymization of author identifiers as implemented in this study, and governance review prior to use in any context involving identifiable individuals or protected groups.

Acknowledgments

The authors acknowledge the computing resources and data access support provided by [Institution Name]. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors unless otherwise stated.

REFERENCES

1. Kleinberg, J. Bursty and hierarchical structure in streams. Data Mining and Knowledge Discovery, 7(4), 373–397.
2. Blei, D. M., Ng, A. Y., & Jordan, M. I. Latent Dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.
3. Lundberg, S. M., & Lee, S. I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
4. Ribeiro, M. T., Singh, S., & Guestrin, C. 'Why should I trust you?': Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
5. Chen, T., & Guestrin, C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
6. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.
7. Hochreiter, S., & Schmidhuber, J. Long short-term memory. Neural Computation, 9(8), 1735–1780.
8. Taylor, S. J., & Letham, B. Forecasting at scale. The American Statistician, 72(1), 37–45.
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
10. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT, 4171–4186.
11. Reimers, N., & Gurevych, I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, 3982–3992.
12. Campello, R. J. G. B., Moulavi, D., & Sander, J. Density-based clustering based on hierarchical density estimates. Pacific-Asia Conference on Knowledge Discovery and Data Mining, 160–172.
13. Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. Comparing Twitter and traditional media using topic models. Advances in Information Retrieval, 338–349.
14. Vosoughi, S., Roy, D., & Aral, S. The spread of true and false news online. Science, 359(6380), 1146–1151.
15. Kwak, H., Lee, C., Park, H., & Moon, S. What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, 591–600.
16. Zhou, X., & Zafarani, R. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys, 53(5), 1–40.
17. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.
18. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. Detection and resolution of rumours in social media: A survey. ACM Computing Surveys, 51(2), 1–36.
19. Castillo, C., Mendoza, M., & Poblete, B. Information credibility on Twitter. Proceedings of the 20th International Conference on World Wide Web, 675–684.
20. Myers, S. A., Zhu, C., & Leskovec, J. Information diffusion and external influence in networks. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 33–41.
21. Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., & Leskovec, J. Can cascades be predicted? Proceedings of the 23rd International Conference on World Wide Web, 925–936.
22. Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. Everyone's an influencer: Quantifying influence on Twitter. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 65–74.
23. Goel, S., Watts, D. J., & Goldstein, D. G. The structure of online diffusion networks. Proceedings of the 13th ACM Conference on Electronic Commerce, 623–638.
24. Yang, J., & Leskovec, J. Patterns of temporal variation in online media. Proceedings of the 4th ACM International Conference on Web

Search and Data Mining, 177–186.
25. Mathioudakis, M., & Koudas, N. TwitterMonitor: Trend detection over the Twitter stream. Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, 1155–1158.
26. Petrović, S., Osborne, M., & Lavrenko, V. Streaming first story detection with application to Twitter. Proceedings of NAACL-HLT, 181–189.
27. Naaman, M., Becker, H., & Gravano, L. Hip and trendy: Characterizing emerging trends on Twitter. Journal of the American Society for Information Science and Technology, 62(5), 902–918.
28. Doshi-Velez, F., & Kim, B. Towards a rigorous science of interpretable machine learning. arXiv preprint, Harvard University Working Paper.
29. Molnar, C. Interpretable machine learning: A guide for making black box models explainable. Leanpub, 2nd edition
30. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215