

BrandShield AI: A Literature Survey on Real-Time Brand-Risk Monitoring Using Sentiment Analysis and Anomaly Detection

Shreyas Nikam¹, Vedant Patil², G. R. Jagtap³

Student (T.Y.), Department of Computer Engineering^{1,2}

Head of the Department, Department of Computer Engineering³

Guru Gobind Singh Polytechnic, Nashik, Maharashtra, India

Abstract: Customer opinions posted online can change the public image of a brand within hours, yet manual monitoring cannot keep pace with the volume and speed of social data. This paper surveys the techniques needed to build an automated, real-time brand-risk monitoring system: lexicon-based, classical, deep, and transformer-based sentiment analysis; aspect-based sentiment analysis; anomaly detection with Isolation Forest; time-series forecasting; information-diffusion modelling; and streaming and dashboard infrastructure. The reviewed studies show that each capability is mature in isolation but that few works combine sentiment, aspect, anomaly, and forecast signals into a single, explainable alert decision, and that regional-language support, noisy text, and evaluation of early crisis detection remain open. Motivated by these gaps, the paper proposes BrandShield AI, a modular pipeline built on DistilBERT, Isolation Forest, and Prophet, and presents its architecture, an alert-fusion rule, and an evaluation plan. The contribution is system-level integration of established methods rather than a new algorithm.

Keywords: anomaly detection, aspect-based sentiment analysis, brand reputation, DistilBERT, Isolation Forest, natural language processing, real-time monitoring, sentiment analysis

I. INTRODUCTION

Online platforms have become the primary place where customers voice satisfaction, frustration, and expectations about products and services. Opinions posted on microblogs, photo-sharing networks, discussion forums, and review portals travel quickly and are visible to a very large audience, so the collective tone of this discussion directly shapes how a brand is perceived [1], [2]. A single widely shared complaint can change that perception within hours, and an organization that learns about the problem only after it has escalated has lost the opportunity to respond early.

Manual observation of brand mentions does not scale to the volume and velocity of modern social data. It is slow, depends on individual judgement, and cannot reliably separate an ordinary fluctuation in opinion from the beginning of a reputational crisis. Commercial social-listening suites address part of this problem but are often priced beyond the reach of small and medium enterprises, and many offer limited support for regional languages and for informal, code-mixed text. Automated methods grounded in Natural Language Processing (NLP) and machine learning have therefore attracted sustained research interest [3], [4], [5].

The research relevant to this problem is broad. Sentiment analysis determines the polarity of a text; aspect-based sentiment analysis (ABSA) attributes that polarity to specific topics; transformer language models supply contextual understanding; anomaly-detection algorithms flag unusual behaviour in a data stream; time-series models forecast future trends; and streaming and dashboard technologies make results usable in real time. These strands are usually studied separately. This paper surveys them from the viewpoint of a single application, *BrandShield AI*, a proposed real-time brand-risk and sentiment-intelligence system, and identifies where an integrated design can add value.



The principal contributions of this survey are as follows:

- a structured review of sentiment-analysis methods, from lexicon-based rules to compact transformers, with emphasis on suitability for streaming social-media text;
- a review of ABSA, anomaly-detection, forecasting, diffusion-modelling, and crisis-detection studies, summarised in comparative tables;
- an explicit statement of the research gaps that motivate an integrated monitoring pipeline; and
- a conceptual architecture, alert-decision formulation, and evaluation plan for BrandShield AI.

The remainder of the paper is organized as follows. Section II describes the survey scope and method. Sections III to VIII review the literature by theme. Section IX compares the studies, Section X states the research gaps, Section XI presents the proposed system, Section XII discusses limitations and ethics, Section XIII outlines future work, and Section XIV concludes.

II. SURVEY SCOPE AND METHOD

This is a narrative literature survey rather than a formal systematic review. Candidate studies were identified by searching the ACL Anthology, arXiv, IEEE Xplore, the ACM Digital Library, and Google Scholar using combinations of the terms *brand reputation*, *social-media sentiment*, *aspect-based sentiment analysis*, *DistilBERT*, *anomaly detection*, *Isolation Forest*, *crisis detection*, *rumour propagation*, and *real-time dashboard*. Foundational works were retained where a technique used by BrandShield AI originates in them, and recent works were retained where they bear directly on brand monitoring.

A study was included when it satisfied at least one of the following criteria: (i) it introduces or evaluates a method for sentiment, aspect, anomaly, forecasting, or diffusion analysis that can be applied to short informal text or time-stamped streams; (ii) it reports a real-time or dashboard-oriented monitoring system; or (iii) it documents a resource, such as a regional-language dataset, that determines the feasibility of a planned extension. Studies were grouped into five themes, shown in Fig. 1: sentiment classification, aspect-level analysis, anomaly detection and forecasting, information diffusion and crisis detection, and streaming and visualization.

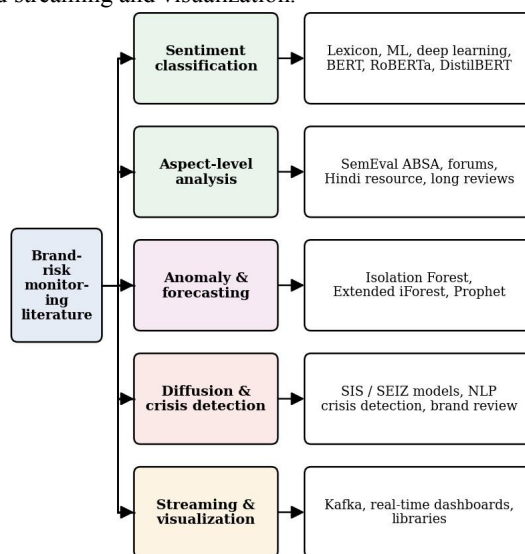


Fig. 1. Taxonomy of the reviewed literature.

Performance figures quoted in this paper are those reported by the cited authors on their own datasets. They are not directly comparable across studies and are not claims about BrandShield AI, whose evaluation is described in Section XI-E.



III. SENTIMENT ANALYSIS APPROACHES

Fig. 2 summarizes how sentiment-analysis methods have progressed. Each stage adds contextual capability, usually at the price of more data and computation, which is the central trade-off for a streaming monitor.

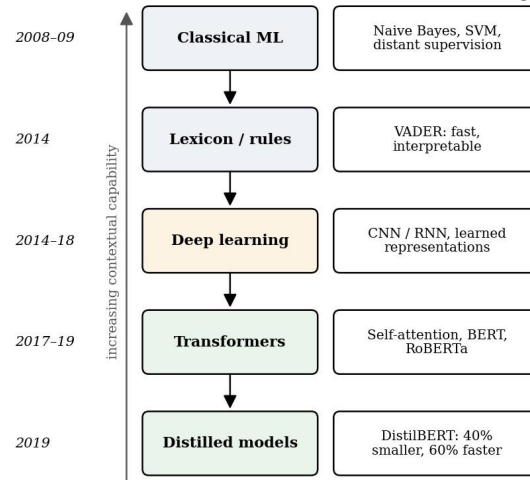


Fig. 2. Progression of sentiment-analysis methods.

A. Lexicon- and Rule-Based Methods

Lexicon-based methods assign a sentiment score to words using a pre-built dictionary and combine the scores with linguistic rules. VADER is the best-known example for social media: it pairs a lexicon tuned to informal text with rules for negation, intensifiers, capitalization, and punctuation, and it requires no training data [6]. Its speed and transparency make it attractive for high-throughput streams. Its weakness is limited contextual depth: a word-level lexicon cannot resolve sarcasm, domain-specific meanings, or sentiment that depends on the surrounding sentence.

B. Classical Machine Learning

Supervised classifiers such as Naive Bayes and support vector machines operate on hand-crafted features, for example n-grams and part-of-speech tags, and were the standard approach in early opinion-mining work [1], [2]. A practical obstacle for social media is the cost of labelled data. Go *et al.* addressed this with distant supervision, using emoticons in posts as noisy labels to train classifiers without manual annotation [7]. Classical models are inexpensive to train but depend heavily on feature engineering and transfer poorly across domains [3].

C. Deep Learning

Recurrent and convolutional networks learn representations directly from word embeddings and removed much of the manual feature engineering. Surveys of this period report consistent gains over classical models on sentence- and document-level sentiment tasks, together with a need for larger training sets and greater computational resources [4]. Recurrent architectures also process tokens sequentially, which limits parallelism and makes very long dependencies difficult to capture.

D. Transformer Language Models

The Transformer architecture replaced recurrence with self-attention, allowing every token to attend to every other token and enabling highly parallel training [8]. BERT pre-trains a bidirectional Transformer on large unlabelled corpora and is then fine-tuned for a downstream task, which produced large improvements across many NLP benchmarks [9]. RoBERTa showed that the same architecture performs better when pre-training is tuned, for example through longer training, more data, and the removal of the next-sentence objective [10].

Because full-size models are costly to serve, compact variants are important for real-time systems. DistilBERT applies knowledge distillation during pre-training to obtain a model that the authors report as 40% smaller and 60% faster than BERT while retaining about 97% of its language-understanding performance [11]. This trade-off is the main reason it is chosen for BrandShield AI.



E. Comparative Evaluations and Domain Adaptation

Joshya compares BERT, DistilBERT, and RoBERTa for sentiment classification on tweet and review datasets and reports that BERT attains the highest accuracy while DistilBERT is faster and more efficient [12]. The finding supports a design choice that favours throughput over the last increment of accuracy in a monitoring setting. TweetEval provides a unified benchmark of tweet-classification tasks, including sentiment, that is useful for evaluating models on the register of microblog text rather than on formal prose [13]. For multilingual deployments, cross-lingual pre-trained encoders such as XLM-R extend transformer-based classification to many languages from a single model [14].

TABLE I
COMPARISON OF SENTIMENT-ANALYSIS PARADIGMS

Paradigm	Strength	Limitation
Lexicon / rules [6]	Fast, interpretable, no training data	Weak on context, sarcasm, domain terms
Classical ML [7]	Low cost; works with distant labels	Feature engineering; poor transfer
Deep learning [4]	Learned representations	Needs more data and compute
Full transformer [9], [10]	Highest contextual accuracy	Heavy for streaming use
Distilled transformer [11]	Compact, fast, strong context	Some accuracy trade-off

IV. ASPECT-BASED SENTIMENT ANALYSIS

A single polarity label for a whole comment hides the information a brand most needs. A review such as one praising a product but criticizing delivery contains opposite opinions about two different aspects. ABSA addresses this by identifying the aspect being discussed and the sentiment expressed toward it. The SemEval-2016 shared task formalized the problem as aspect-category detection, opinion-target extraction, and aspect-level polarity classification over customer reviews in several domains and languages [15].

Fig. 3 illustrates the idea with a constructed example: one comment yields a positive opinion about the product and negative opinions about delivery and service, and aggregating these by time window shows which concern is growing.

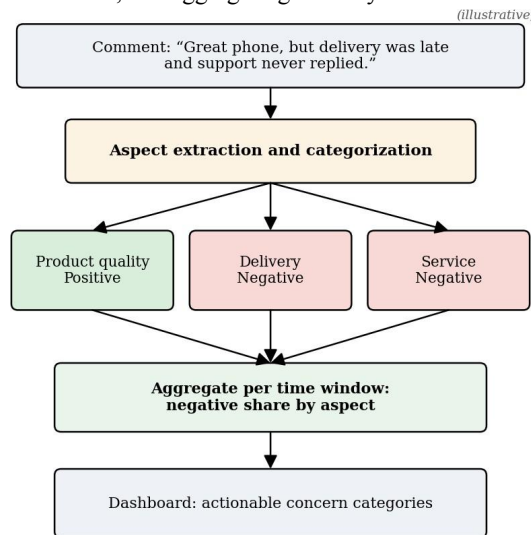


Fig. 3. Aspect-based analysis flow (constructed example).

Participating systems illustrate the main design options. SIEL, submitted to SemEval-2015, performed ABSA on reviews by combining aspect identification with sentiment assignment [16]. AKTSKI, submitted to SemEval-2016, tackled consumer-review ABSA and showed how opinions can be tied to specific product or service aspects [17]. Pateria later examined how sentiment flows across a sentence, using both nearby and non-local context to predict aspect-level polarity, which is relevant when one comment carries several, sometimes contrasting, opinions [18].



Two further lines of work extend the reach of ABSA. Zhang *et al.* applied it to question-answering forums, showing that aspect and sentiment extraction transfer beyond conventional product reviews [19]. This supports a broader input strategy in which BrandShield AI may eventually process social posts, reviews, news items, and forum text. Akhtar *et al.* created an annotated Hindi resource for aspect-based analysis of product reviews [20], which is directly relevant to the Indian-language extension considered in this project. Most recently, E-ABSA20K introduced a 20,000-review multi-domain dataset and a propose-and-verify approach for long e-commerce reviews with many aspects [21], which anticipates a challenge that real-world brand comments are likely to pose.

For BrandShield AI the relevance is practical. Negative comments organized by aspect, for example product quality, service, price, or delivery, are actionable, whereas an undifferentiated count of negative posts is not. None of the reviewed ABSA studies, however, connects aspect output to temporal anomaly detection or to an alerting layer.

V. ANOMALY DETECTION FOR REPUTATION MONITORING

A. Taxonomy

Chandola *et al.* classify anomalies as point anomalies (an individual observation that deviates from the rest), contextual anomalies (an observation that is unusual only in a particular context such as time of day), and collective anomalies (a group of observations that is unusual together) [22]. The distinction matters for brand monitoring. A burst of negative posts at three in the morning may be unremarkable for a global brand and alarming for a local one, and a campaign launch may legitimately produce a surge that is not a crisis. Treating every deviation as an incident would generate false alerts.

B. Isolation Forest

Isolation Forest detects anomalies by isolation rather than by profiling normal behaviour. It builds an ensemble of random trees by repeatedly choosing a random feature and a random split value; anomalous points, being few and different, tend to be isolated after fewer splits, so a short average path length indicates an anomaly [23]. For a sample of size n , the anomaly score of an observation x is

$$s(x, n) = 2^{-E[h(x)]/c(n)} \quad (1)$$

where $h(x)$ is the path length of x in a tree, $E[h(x)]$ is its mean over the forest, and the normalizing term is

$$c(n) = 2H(n-1) - 2(n-1)/n(2)$$

with $H(i)$ the harmonic number, approximated by $\ln(i) + 0.5772$. Scores near 1 indicate likely anomalies, and scores well below 0.5 indicate normal observations [23]. The method has low computational cost and scales to large datasets, properties that suit streaming use. Hariri *et al.* later proposed Extended Isolation Forest, which replaces axis-parallel splits with randomly oriented hyperplanes and removes score artefacts that axis-aligned splits can introduce [24]. Open-source implementations are available in standard machine-learning libraries [25].

C. Applying Anomaly Detection to Sentiment Streams

In a sentiment-monitoring pipeline the detector does not operate on raw text. Posts are first classified, then aggregated per time window into features such as the negative-sentiment ratio, mention volume, and the growth rate of each. The anomaly detector then scores each window. Two cautions follow from the literature. First, Isolation Forest identifies statistical rarity; it does not explain the cause, nor does it establish that the event is harmful to the brand [23], [22]. Second, a fixed threshold is brittle because baselines differ between brands and vary with time. Adaptive thresholds and fusion with independent signals, for example sentiment polarity and trend direction, are therefore needed to separate genuine risk from normal variation.

Fig. 4 illustrates both ideas on synthetic data: a window-level negative ratio rising above its rolling baseline, and the isolation principle by which a rare point is separated from a dense cluster in few random splits.



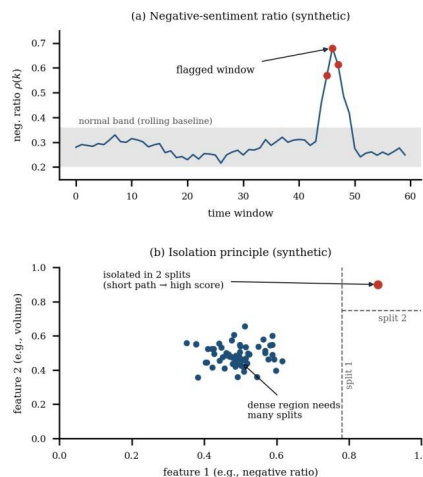


Fig. 4. Anomaly detection illustrated: (a) a flagged window; (b) isolation by random splits. Data are synthetic.

VI. TIME-SERIES FORECASTING OF SENTIMENT

Prophet is a forecasting procedure designed for business time series with strong seasonal effects and an analyst in the loop. It uses a decomposable additive model

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (3)$$

in which $g(t)$ is a piecewise trend, $s(t)$ captures periodic seasonality, $h(t)$ models holiday or event effects, and the final term is residual error [26]. The approach tolerates missing values and outliers and exposes interpretable parameters that analysts can adjust, which makes it suitable for a small team without specialist forecasting skills. For BrandShield AI, forecasting the negative-sentiment ratio supports a forward-looking signal: a sustained upward trajectory can raise concern before a hard anomaly occurs. The original work focuses on forecasting itself and does not address text analytics, live social-media ingestion, or crisis detection [26], so its use here is as one component of a larger pipeline.

VII. INFORMATION DIFFUSION AND CRISIS DETECTION

A. Epidemic Models of Information Spread

Jin *et al.* modelled how news and rumours propagate on Twitter using epidemiological compartment models. Treating exposure to a story like exposure to a disease, they compared the classical susceptible-infected-susceptible (SIS) model with the SEIZ model, which adds exposed and sceptic compartments, fitted the models with differential equations, and found SEIZ to describe the observed propagation better; the models also helped distinguish rumours from genuine news [27]. The study does not use NLP or sentiment analysis, and it is not a real-time system. Its relevance is conceptual: it supports the view that negative content spreads with dynamics that can be characterized and that early growth is informative.

B. NLP-Based Crisis Detection Systems

Nithya *et al.* describe a real-time crisis-detection approach for social media that combines NLP preprocessing (tokenization, stemming, lemmatization), deep models such as BERT and LSTM, named-entity and embedding features, sentiment analysis, and Apache Kafka for streaming, with a dashboard for decision support [28]. As reported, real-time operation was associated with improved crisis awareness ($r = 0.61$) and a smaller association with decision-making ($r = 0.43$). The reported limitations include weak multilingual handling and limited discussion of real-world deployment.

C. Sentiment for Brand Reputation Management

Brooklyn *et al.* review sentiment analysis for brand-reputation management, covering lexicon-based, supervised (for example SVM and Naive Bayes), hybrid, and deep-learning (RNN and Transformer) models. They conclude that real-



time sentiment tracking helps detect reputation problems early and supports customer engagement and decisions, while noting persistent difficulty with sarcasm and context, noise from spam and fake content, and limited crisis-prediction capability [29].

VIII. STREAMING INFRASTRUCTURE AND DASHBOARDS

Real-time monitoring requires a transport that decouples data producers from processing stages. Kafka is a distributed publish-subscribe log designed for high-throughput collection and delivery of event data, allowing several consumers to process the same stream at their own rate [30]. In a pipeline, this permits ingestion, preprocessing, and model inference to scale independently and to continue operating if one stage is slow.

Pre-trained transformer models can be loaded and served through widely used open-source libraries [31], which lowers the engineering effort for a small team. The study that is closest to BrandShield AI is the real-time dashboard of Reddy *et al.*, which collects posts from X, preprocesses them, scores them with VADER and DistilBERT, and visualizes the results for corporate-reputation monitoring; the authors report an accuracy of about 87% for the hybrid approach [32]. The work shows that combining a fast lexicon method with a contextual transformer is practical, but it stops at sentiment visualization: it has no anomaly or crisis detector, no aspect analysis, and no forecasting.

IX. COMPARATIVE ANALYSIS OF REVIEWED STUDIES

Tables II and III summarize the reviewed work by technique, contribution, and limitation relative to an integrated brand-monitoring objective.

TABLE II
SENTIMENT AND ASPECT-LEVEL STUDIES

Study	Technique	Main contribution	Gap for brand monitoring
[12]	BERT / DistilBERT / RoBERTa	Accuracy-speed comparison	No alerts or dashboard
[16]	ABSA	Aspect-level review analysis	Not a monitoring platform
[17]	ABSA	Consumer-review aspects	No risk-alert layer
[18]	Context-aware ABSA	Local and non-local context	Prediction only
[19]	ABSA on forums	Broader input domains	No anomaly or dashboard
[20]	Hindi ABSA resource	Indian-language dataset	Language scope only
[21]	Long-review ABSA	Multi-aspect long reviews	Beyond prototype scope

TABLE III
ANOMALY, FORECASTING, AND SYSTEM STUDIES

Study	Technique	Main contribution	Gap for brand monitoring
[23]	Isolation Forest	Efficient isolation of anomalies	No semantic explanation
[24]	Extended iForest	Removes split-bias artefacts	Generic, not text-specific
[26]	Additive forecasting	Trend, seasonality, holidays	No NLP or crisis logic
[27]	SEIZ / SIS models	Rumour propagation dynamics	No NLP, not real-time
[28]	BERT/LSTM + Kafka	Real-time crisis awareness	Multilingual, deployment
[29]	ML/DL sentiment review	Brand-reputation synthesis	Sarcasm, noise, prediction
[32]	VADER + DistilBERT	Real-time sentiment dashboard	No anomaly/aspect/forecast

X. RESEARCH GAPS

The comparative analysis shows that each building block is mature in isolation. The gaps lie in how the blocks are combined and validated for brand monitoring:



1. **Fragmented capabilities.** Sentiment, aspect, anomaly, forecasting, and alerting are rarely combined in one practical workflow. The closest real-time dashboard stops at sentiment display [32].
2. **Anomalies without context.** Isolation Forest flags rarity but not cause or business significance [23]. Few brand-monitoring studies fuse anomaly scores with polarity, aspect, and trend signals to limit false alerts.
3. **Reactive rather than predictive.** Most monitoring work reports present sentiment; forecasting is studied separately from text analytics [26], [29].
4. **Language coverage.** Regional-language and code-mixed text remain under-served in monitoring systems [20], [28], although resources and multilingual encoders exist [14].
5. **Informal and noisy text.** Sarcasm, slang, spam, and fake content continue to degrade accuracy [29].
6. **Accessibility and cost.** Proprietary suites are expensive; open, modular, low-cost designs aimed at small organizations are under-documented.
7. **Evaluation.** There is no common benchmark for early-crisis detection in brand streams, so system claims are hard to compare. Reported correlations for crisis awareness are moderate [28].

Fig. 5 maps the reviewed studies against the capabilities needed for brand-risk monitoring, using the contributions described in Sections III to VIII. No single reviewed study covers more than three of the seven capabilities, and none combines anomaly detection with aspect analysis or forecasting.

	Sentiment	Aspect	Anomaly	Forecast	Real-time / dashboard	Alerts	Regional language
VADER	✓						
DistilBERT	✓						
BERT-family eval.	✓						
SemEval ABSA	✓	✓					
Hindi ABSA	✓	✓					✓
Isolation Forest			✓				
Prophet				✓			
Reddy et al.	✓				✓		
Nithya et al.	✓				✓	✓	
Brooklyn et al.	✓				✓		
BrandShield AI	✓	✓	✓	✓	✓	✓	planned

Fig. 5. Capability coverage of reviewed studies versus the proposed system.

XI. PROPOSED SYSTEM: BRANDSHIELD AI

BrandShield AI is intended to address gaps 1 to 3 and 6 directly and to prepare for gaps 4 and 5. Its contribution is system-level integration and practical implementation of established techniques, not a new sentiment or anomaly algorithm.

A. Architecture

Fig. 6 shows the processing pipeline. Data from social-media APIs, news feeds, or prepared datasets is published to a Kafka stream, preprocessed, and passed through sentiment, aspect, and trend stages. The resulting window features feed the anomaly detector and the forecaster, whose outputs are fused by an alert engine. Raw and processed data are



stored for analysis and retraining, and results are exposed through REST services to a web dashboard and to notification channels.

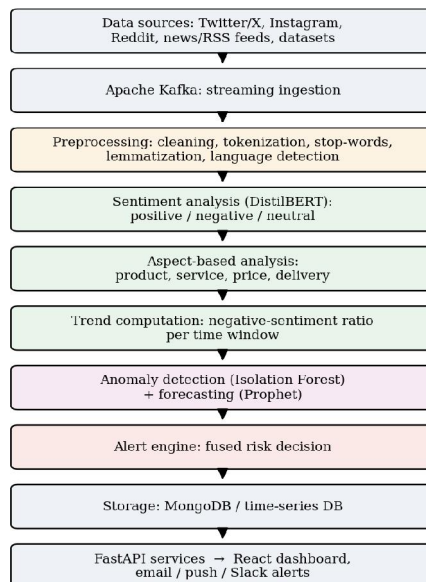


Fig. 6. Conceptual processing pipeline of BrandShield AI.

Fig. 7 shows the corresponding layered deployment view. The React dashboard communicates with a FastAPI backend, which drives the AI pipeline; the pipeline in turn depends on external services and on separate stores for raw data and time-series aggregates.

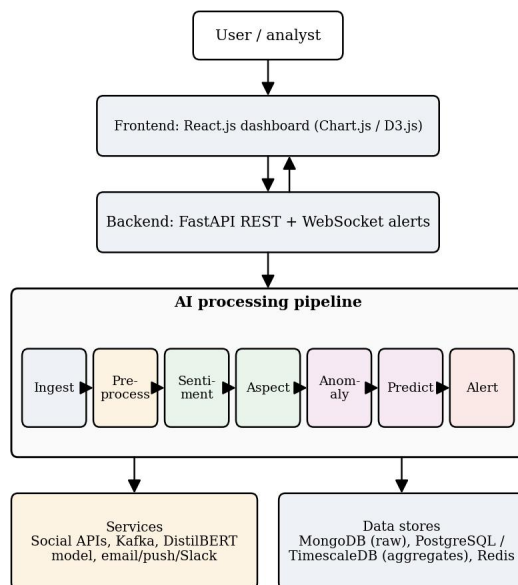


Fig. 7. Layered system architecture of BrandShield AI.

B. Module Design

- **Ingestion and preprocessing.** Posts are collected through permitted APIs or datasets, then cleaned of URLs, mentions, and noise, tokenized, stripped of stop-words, lemmatized, and tagged by language.



- **Sentiment.** A fine-tuned DistilBERT classifier labels each post positive, negative, or neutral [11]. A lexicon baseline [6] is retained for comparison.
- **Aspect analysis.** Posts are mapped to concern categories (product quality, service, price, delivery) so that negative volume can be broken down by cause.
- **Anomaly detection.** An Isolation Forest scores per-window features [23].
- **Forecasting.** Prophet projects the negative-ratio series forward [26].
- **Dashboard and alerts.** Sentiment share, trends, aspect breakdown, anomalies, forecasts, and alerts are presented in a web interface; email, push, and messaging notifications deliver urgent events.

C. Alert-Decision Formulation

For time window k , let $N(k)$ be the number of posts and $N^-(k)$ the number classified negative. The negative-sentiment ratio is

$$\rho(k) = N^-(k)/N(k) \quad (4)$$

Let $a(k)$ be the Isolation Forest anomaly score of the window feature vector, and let μ and σ be the rolling mean and standard deviation of ρ over a trailing baseline. A risk alert is raised only if all of the following hold:

$$a(k) \geq \tau_a \wedge \rho(k) > \mu + z\sigma \wedge \hat{\rho}(k+1) \geq \rho(k) \quad (5)$$

Here τ_a is the anomaly threshold, z a sensitivity parameter set through user configuration, and $\hat{\rho}(k+1)$ the forecast for the next window. Requiring agreement among rarity, elevated negativity, and an upward forecast operationalizes the fusion recommended in Section V-C and is intended to reduce false alerts. This is a design proposal; its thresholds will be tuned empirically and are not yet validated.

Fig. 8 expresses the rule as a decision flow. Each failed test routes the window to a lower-severity outcome, so that only windows passing all three tests reach the alert stage.

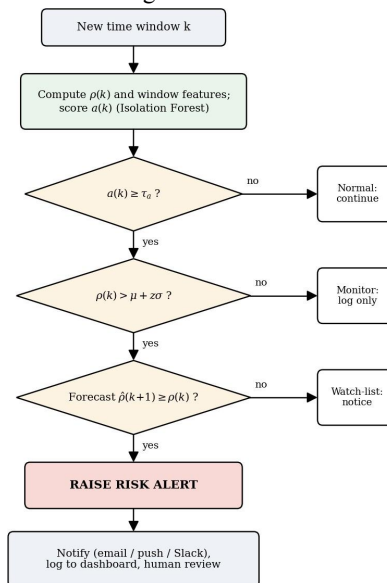


Fig. 8. Alert-decision flowchart for each time window.



D. Technology Stack and Implementation Challenges

TABLE IV
PLANNED TECHNOLOGY STACK

Layer	Technologies
Frontend	React.js / Next.js, Tailwind CSS, Chart.js / D3.js
Backend	FastAPI / Flask, REST APIs, WebSockets
ML / NLP	DistilBERT, Isolation Forest, Prophet, NLTK, spaCy
Streaming	Apache Kafka
Storage	MongoDB, PostgreSQL / TimescaleDB, Redis cache
Deployment	Docker, GitHub Actions, AWS / GCP

TABLE V
ANTICIPATED CHALLENGES AND MITIGATION

Challenge	Planned mitigation
Noisy social data	NLP preprocessing; spam and noise filtering
Contextual sentiment	DistilBERT with domain-specific validation
Crisis vs. normal fluctuation	Isolation Forest with adaptive thresholds
Real-time processing	Kafka with asynchronous processing
False crisis alerts	Fuse sentiment, anomaly, and trend signals

E. Evaluation Plan

The system will be evaluated on a labelled or prepared dataset and, where permitted, on live API data. Sentiment quality will be reported as accuracy and macro-averaged F1 against the lexicon baseline. Anomaly and alert quality will be reported as precision, recall, and false-alert rate against windows annotated or injected as crisis events, together with detection lead time relative to a manual-monitoring baseline. System behaviour will be reported as end-to-end latency and throughput. No experimental results are reported in this paper.

XII. LIMITATIONS AND ETHICAL CONSIDERATIONS

Technical. Model accuracy depends on data quality and volume and will degrade on sarcasm, slang, and mixed-language text. A distilled model trades some accuracy for speed [11], [12]. Isolation Forest scores indicate rarity only, so alert quality depends on the fusion rule and on well-chosen thresholds. Forecasting accuracy depends on having enough history to estimate seasonality.

Data access. Real-time monitoring depends on the availability, cost, and terms of third-party APIs, which can change without notice. The prototype therefore supports prepared datasets as well as permitted live sources.

Ethical. Social posts are personal data. A deployed system should use only permitted sources, minimize and anonymize stored identifiers, apply retention limits, and avoid profiling individual users. Classifiers may inherit bias from training data, so outputs should support rather than replace human judgement, and flagged events should be reviewed by a person before any public response.

XIII. FUTURE RESEARCH DIRECTIONS

- Multilingual and code-mixed processing using cross-lingual encoders [14] and regional resources [20].
- Stronger aspect models, including handling of long, multi-aspect comments [21].
- Spam, fake-review, and duplicate-content detection to reduce noise.
- Extended Isolation Forest [24] and diffusion-aware features inspired by epidemic models [27].
- Multimodal analysis of images and video, and scalable cloud deployment.
- A shared benchmark and annotation protocol for early brand-crisis detection.



XIV. CONCLUSION

This paper surveyed the techniques that underpin real-time brand-risk monitoring. Lexicon methods offer speed and transparency; compact transformers such as DistilBERT offer contextual understanding at modest cost; aspect-based analysis explains what customers are unhappy about; Isolation Forest provides an efficient detector of unusual behaviour; Prophet offers interpretable forecasts; and streaming and dashboard technologies make results available in time to act. Existing studies treat these capabilities mostly as separate components, and the closest real-time dashboards stop at sentiment display.

BrandShield AI is proposed to close this gap through an integrated, low-cost pipeline in which sentiment, aspect, anomaly, and forecast signals are fused into a single alert decision. The contribution is practical system integration of established methods rather than a new algorithm. Implementation and empirical evaluation, following the plan in Section XI-E, are the next steps.

REFERENCES

1. B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.
2. B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, 2012.
3. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
4. L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Min. Knowl. Discov.*, vol. 8, no. 4, Art. no. e1253, 2018.
5. J. Wong, "Sentiment analysis on social media: A survey," 2025. [Online]. (Bibliographic details to be verified against the source copy.)
6. C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. AAAI Conf. Web Social Media (ICWSM)*, 2014, pp. 216–225.
7. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Rep., Stanford Univ., Stanford, CA, USA, 2009.
8. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017.
9. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
10. Y. Liu *et al.*, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv:1907.11692, 2019.
11. V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv:1910.01108, 2019.
12. Joshy, "Performance comparison of BERT, DistilBERT, and RoBERTa for sentiment analysis," 2022. (Bibliographic details to be verified against the source copy.)
13. F. Barbieri, J. Camacho-Collados, L. Espinosa-Anke, and L. Neves, "TweetEval: Unified benchmark and comparative evaluation for tweet classification," in *Findings Assoc. Comput. Linguistics: EMNLP*, 2020, pp. 1644–1650.
14. Conneau *et al.*, "Unsupervised cross-lingual representation learning at scale," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2020, pp. 8440–8451.
15. M. Pontiki *et al.*, "SemEval-2016 Task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 19–30.
16. S. Guha, A. Joshi, and V. Varma, "SIEL: Aspect based sentiment analysis in reviews," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 759–766.
17. S. Pateria and P. Choubey, "AKTSKI at SemEval-2016 Task 5: Aspect based sentiment analysis for consumer reviews," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 318–324.



18. S. Pateria, "Aspect based sentiment analysis using sentiment flow with local and non-local neighbor information," in *Proc. COLING*, 2016, pp. 2635–2646.
19. W. Zhang, Y. Deng, X. Li, L. Bing, and W. Lam, "Aspect-based sentiment analysis in question answering forums," in *Findings Assoc. Comput. Linguistics: EMNLP*, 2021, pp. 4582–4591.
20. M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect based sentiment analysis in Hindi: Resource creation and evaluation," in *Proc. 10th Int. Conf. Lang. Resources Eval. (LREC)*, 2016.
21. T. Sun, M. Ma, and C. Yu, "E-ABSA20K: A dataset and propose-and-verify for aspect-based sentiment analysis in long e-commerce reviews," in *Findings Assoc. Comput. Linguistics: ACL*, 2026, pp. 17959–17973.
22. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Art. no. 15, 2009.
23. F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining (ICDM)*, 2008, pp. 413–422.
24. S. Hariri, M. C. Kind, and R. J. Brunner, "Extended isolation forest," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1479–1489, Apr. 2021.
25. F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
26. S. J. Taylor and B. Letham, "Forecasting at scale," *Amer. Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
27. F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on Twitter," in *Proc. 7th Workshop Social Netw. Mining Anal. (SNAKDD)*, 2013, Art. no. 8.
28. Nithya *et al.*, "Real-time crisis detection on social media using NLP," 2024. (Bibliographic details to be verified against the source copy.)
29. P. Brooklyn *et al.*, "Social media sentiment analysis for brand reputation management," 2024. (Bibliographic details to be verified against the source copy.)
30. J. Kreps, N. Narkhede, and J. Rao, "Kafka: A distributed messaging system for log processing," in *Proc. NetDB Workshop*, 2011.
31. T. Wolf *et al.*, "Transformers: State-of-the-art natural language processing," in *Proc. EMNLP: Syst. Demonstrations*, 2020, pp. 38–45.
32. Y. A. Reddy, S. Agarwal, V. Parashar, and A. Arora, "Visualizing public opinion on X: A real-time sentiment dashboard using VADER and DistilBERT," arXiv:2504.15448, 2025.

