NLP NEWS INTELLIGENCE ANOMALY RADAR

Presenter

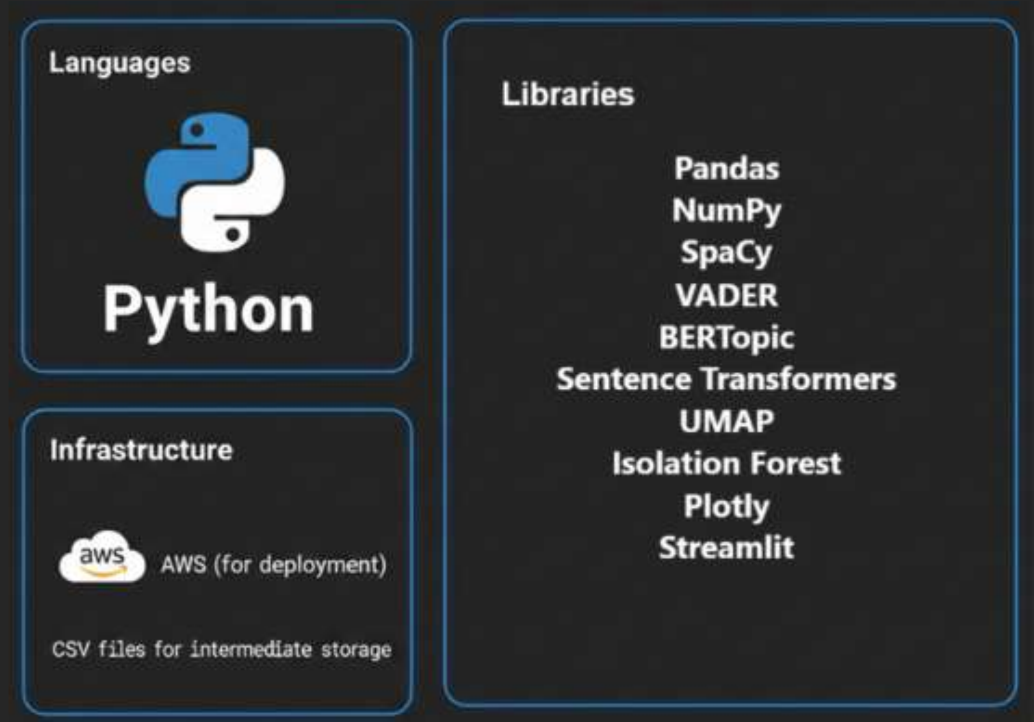
Geetha Palanisamy
MDTM46B

PROJECT OVERVIEW

PROJECT GOAL

- Build a platform to detect suspicious narratives, monitor trends, and assess brand risk from news articles.
- This provide analysts with early warning signals of disinformation and reputational exposure.

TOOLS USED



Project Structure

Step 1: Data ingestion → raw news dataset

Step 2: Preprocessing → cleaning text, normalizing locations

Step 3: Feature engineering → sentiment, topics, temporal signals

Step 4: Anomaly detection → linguistic, location, temporal

Step 5: Aggregation → article labels, brand risk scores

Step 6: Visualization → UMAP clusters, dashboards

Step 7: Deployment → Streamlit, AWS

Why chose Unsupervised model?

- Unlike supervised detection that only recognize known patterns of fake news, unsupervised isolation forest can detect Zero-day disinformation
- It overcomes label scarcity
- Supervised models often struggle with extreme class imbalance
- Reduced the need of constant model retraining
- Supervised classifiers miss the mark on unknowns. Our NLP News Intelligence Radar uncovers hidden patterns with isolation forest and Z score thresholding staying ahead of evolving disinformation that rigid models ignore

The problem(supervised)	The solution(unsupervised)
• Rigid Patterns	• Zero day detection
• Label Scarcity	• Unsupervised power
• Class imbalance	• Anomaly Mastery
• High maintenance	• Evolving intelligence

DATA HANDLING

Data Cleaning

- Removed URLs, punctuation, and special characters so the text is easier to analyze.
- Converted everything to lowercase for consistency.
- Removed common filler words (like “the”, “and”, “is”) that don’t add meaning.
- Applied lemmatization: turning words into their base form (e.g., “running” → “run”).
- Why: Clean text ensures that our models focus on meaning, not noise.

Location Extraction

1. SpaCy Named Entity Recognition (NER) is a pretrained model to detect entities like GPE (Geo-Political Entity), LOC (location), and ORG. Example: “US oil prices slip below 50” → SpaCy tags “US” as a GPE
2. **GeoText :**A lightweight python library that scans text for city and country names. Example: “England football fans party in Lille” → GeoText extracts “England” and “Lille”.
3. Normalization :After extraction, standardized names (e.g., “US”, “USA”, “United States” → “United States”) as Consistency is critical for anomaly detection — otherwise “US” vs “USA” would look like different places.

FEATURE ENGINEERING

Sentiment Analysis

- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a nltk module used for sentiment analysis
- **Lexicon-based approach:** VADER has a dictionary of words with pre-assigned sentiment scores (positive, negative, neutral). Example: “great” = +3.0, “terrible” = −3.0.
- Why best Approach? : No training needed, Fast and light weight, Transparent and explainable even for short text

Temporal Features

•Extracted **year, month, and weekday** from the article date.

•Counted how many articles appeared per day.

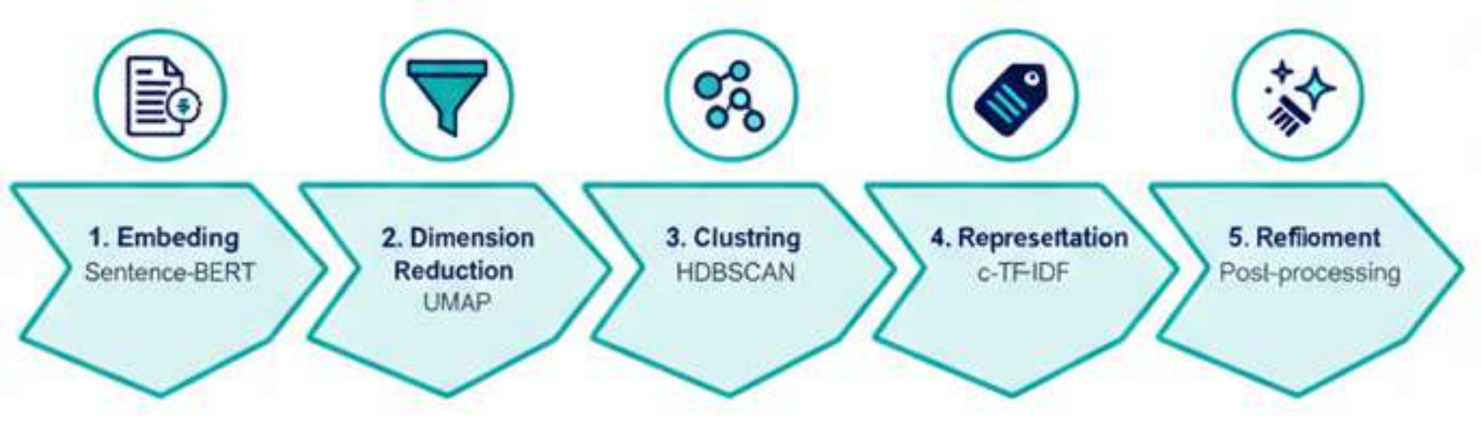
•**Why:** Sudden spikes in volume often indicate coordinated campaigns.

FEATURE ENGINEERING

Topic Extraction – BERTopic

- Why?- capture meaning not just frequency like LDA

- Steps involved

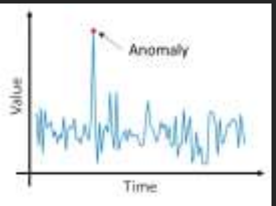


- Sentence transformer read each document and output a vectors of numbers called an embedding (through semantic analysis-similar meaning)
- This vector captures meaning of text in a high dimensional space
- UMAP takes this high dimensional embeddings and compress them into a small vector(2/5D) , but groups similar text close, dissimilar texts far
- Now BERTopic via DBSCAN, clusters these reduced vectors as 0,1,2…until this step there is no text label, all words are in semantic encoding(i.e.numbers)
- Now, BERTopic takes the original text of that cluster and merge them into single document and run its c-TF-IDF to find the most representative words for that cluster (number to words form)
- Top scoring words in that cluster (e.g:refund, delay,shipment) are selected as keywords of that topic
- After seeing these keywords of a cluster human gives short topic name

ANOMALY DETECTION

Temporal Anomaly

- For temporal anomaly detection: **rolling mean + rolling standard deviation with z-score thresholding**.
- **Rolling mean/std:** Calculates the average and variability of article counts over a moving window (e.g., 7 days).
- **Z-score:** Measures how far today's article count deviates from the rolling average in terms of standard deviations.
- **Threshold:** If the z-score exceeds a set value (e.g., >2), it's flagged as a temporal anomaly.



Why temporal features matter?

- **Disinformation campaigns often spike suddenly** — many articles appear in a short time to push a narrative.
- **Normal news cycles are smoother** — gradual coverage over time.
- **Temporal anomalies highlight coordination** — e.g., 50 articles about the same protest appearing overnight.

🔄 Alternatives you could have used:

- **ARIMA / Prophet (time series forecasting):** Predict expected counts, flag deviations. More complex, less transparent.

ANOMALY DETECTION

Linguistic Anomaly

- **Isolation Forest** is an unsupervised machine-learning algorithm based on the idea that anomalies are **rare and different** from normal data points.
- **Features: text length, tone, and topic meaning.**
- **Mechanism:**

Step 1 — Represent each article: Every article is converted into a profile using these features.

Step 2 — Build random trees: The algorithm creates many random decision trees that split articles based on these features.

Step 3 — Isolation principle:

Articles that resemble most others require many splits before they can be separated.

Articles that differ clearly are isolated quickly in fewer splits.

Step 4 — Anomaly scoring: Articles that are consistently isolated faster across many trees receive higher anomaly scores. These are flagged as potential anomalies.

Why isolation forest?

- **It does not require labeled data to train.**
- **It explicitly targets anomalies rather than profiling normal points.**

ANOMALY DETECTION

Location Anomaly

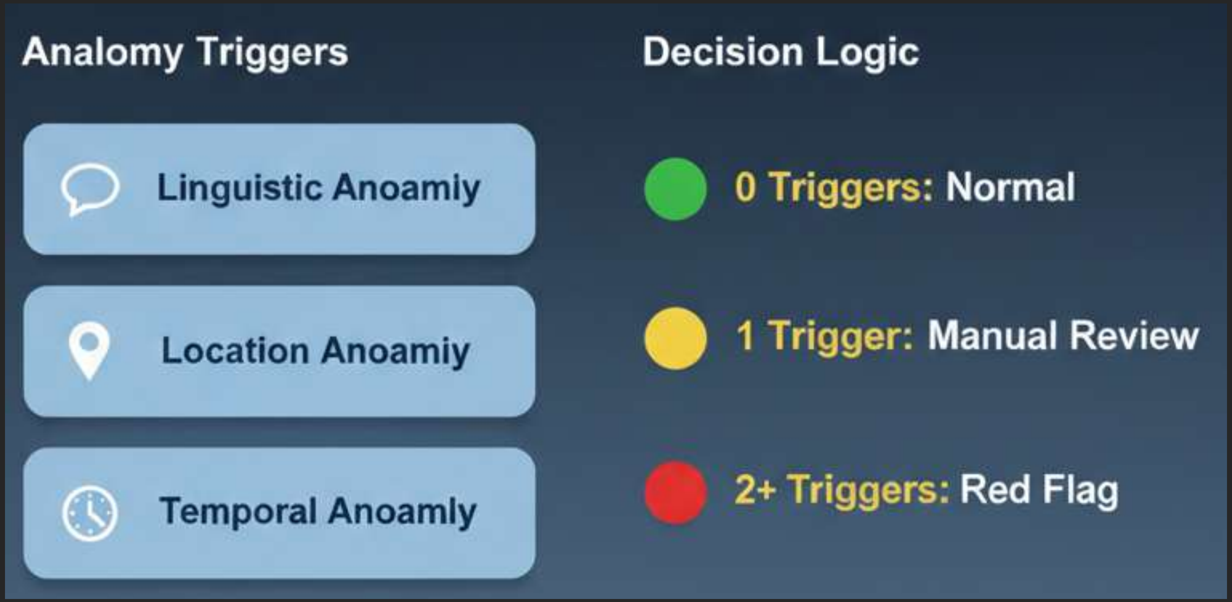
- **en_core_web_sm, a pre-trained English language model provided by the SpaCy library, and it is used to perform Named Entity Recognition (NER)**
- **Rule-based comparison:** Simple string/semantic match between headline location and body location.
- **Binary anomaly flag:** 0 = normal, 1 = anomaly.

Why location anomaly matters

- **Disinformation tactic:** Fake stories often recycle content but swap locations to mislead readers.
- **Explainability:** Easy for analysts to verify — “Headline says London, body says Paris.”
- **Transparency:** Non-technical clients can immediately understand why it’s flagged.
- **Business relevance:** Location mismatches can damage trust in reporting and mislead audiences about where events occur.

AGGREGATION

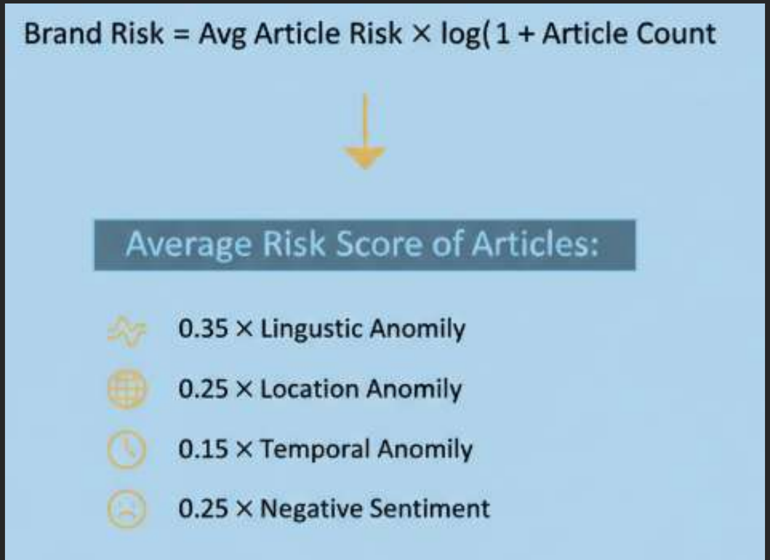
Final Label



Why?

- Transparent and easy
- Provides clear visibility to specific triggers

Brand Risk Score



- 0.35 : Unusual wording is often the clearest sign of manipulation.
- 0.25 : Location mismatches and negative sentiment are equally important — they directly affect trust and reputation.
- 0.15 : volume alone doesn’t always mean risk.

EVALUATION METRICS

How Well Does Our Risk Detection Work?

Metric	Score	Key Takeaway	Insight
ROC_AUC	1	Perfect Discrimination	Our model perfectly distinguishes between suspicious and normal articles.
PR_AUC	1	High Precision Balance	Even when risky articles are rare, we detect them with maximum confidence.
Precision	1	Zero False Alarms	100% of flagged articles are truly risky; no time is wasted on false positives.
F1 Score	0.06	Precision-Heavy	We currently prioritize extreme accuracy over total volume.
Recall@200	24%	Targeted Review	Reviewing the top 200 articles catches nearly a quarter of all known risks.

Key Takeaway

- **This is not a bulk detection tool — it is a decision-support system.** It delivers **accurate, explainable, and actionable alerts**, enabling faster and more confident human review.

Visualization – Tab1

Disinformation Detection

Inference:

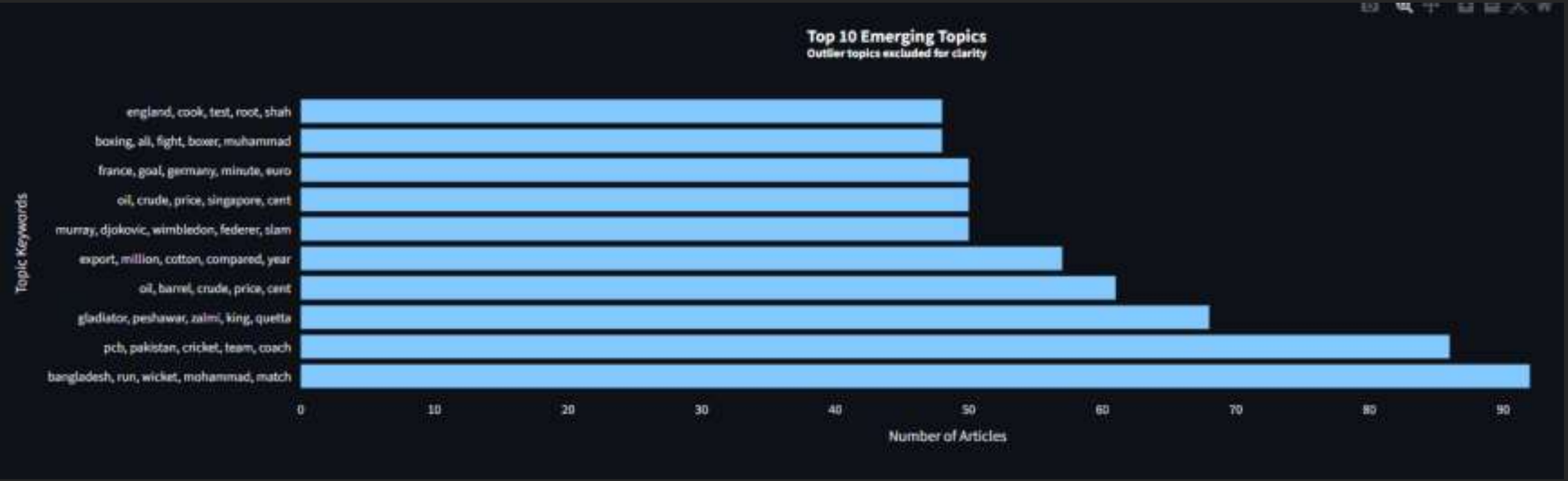
- Similar articles with elevated risk are potentially coordinated narratives
- Same story, different places is classic sign of recycled or misleading content.
- Dense areas indicate strong narrative repetition; sparse areas are unique or outlier stories.

Visualization-Tab2

Hyperlocal Trend monitoring

Sentiment trends:

- If negativity spikes in certain years, it may signal crises, manipulation, or coordinated campaigns.
- If positivity rises, it may reflect recovery or positive narratives.
- Neutral coverage shows stability.

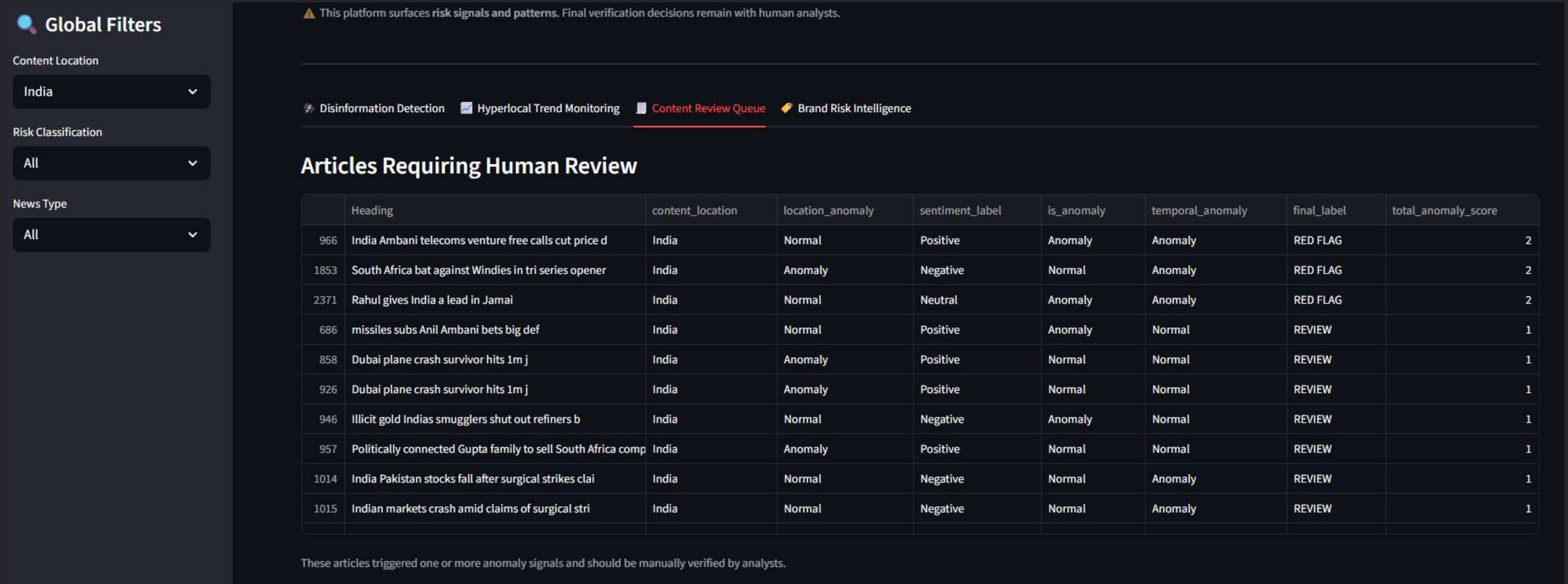


Topic dominance:

- Top topics highlight what the media is focusing on (e.g., energy, politics, health).
- Emerging topics show new narratives gaining attention.

Visualization-Tab3

Content Review



Articles Requiring Human Review

	Heading	content_location	location_anomaly	sentiment_label	is_anomaly	temporal_anomaly	final_label	total_anomaly_score
966	India Ambani telecoms venture free calls cut price d	India	Normal	Positive	Anomaly	Anomaly	RED FLAG	2
1853	South Africa bat against Windies in tri series opener	India	Anomaly	Negative	Normal	Anomaly	RED FLAG	2
2371	Rahul gives India a lead in Jamai	India	Normal	Neutral	Anomaly	Anomaly	RED FLAG	2
686	missiles subs Anil Ambani bets big def	India	Normal	Positive	Anomaly	Normal	REVIEW	1
858	Dubai plane crash survivor hits 1m j	India	Anomaly	Positive	Normal	Normal	REVIEW	1
926	Dubai plane crash survivor hits 1m j	India	Anomaly	Positive	Normal	Normal	REVIEW	1
946	Illicit gold Indias smugglers shut out refiners b	India	Normal	Negative	Anomaly	Normal	REVIEW	1
957	Politically connected Gupta family to sell South Africa comp	India	Anomaly	Positive	Normal	Normal	REVIEW	1
1014	India Pakistan stocks fall after surgical strikes clai	India	Normal	Negative	Normal	Anomaly	REVIEW	1
1015	Indian markets crash amid claims of surgical stri	India	Normal	Negative	Normal	Anomaly	REVIEW	1

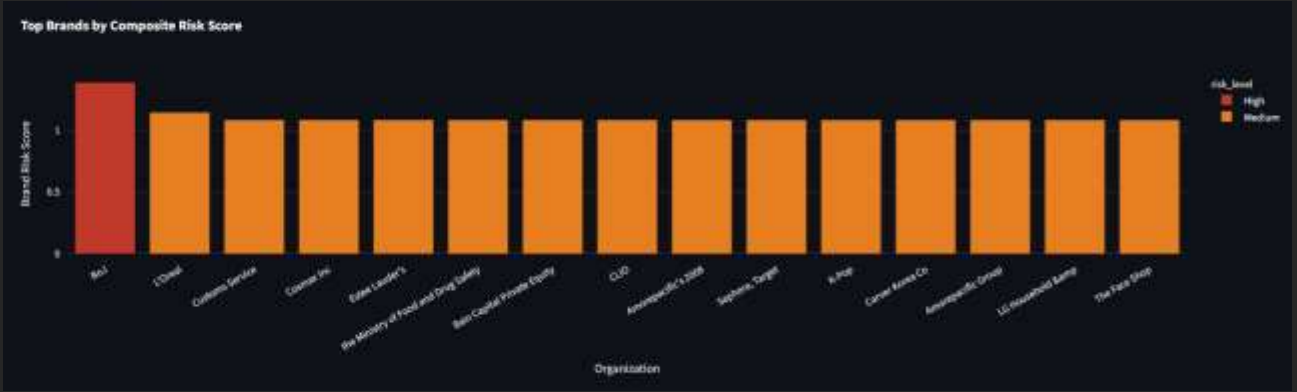
These articles triggered one or more anomaly signals and should be manually verified by analysts.

Visualization-Tab4

Brand Risk

📊 Brand Risk Summary

	organization	avg_article_risk	article_count	brand_risk_score	risk_level
0	BoJ	0.7102	6	1.382	High
1	L'Oreal	0.5861	6	1.1405	Medium
12	Customs Service	0.6028	5	1.08	Medium
10	Cosmax Inc	0.6028	5	1.08	Medium
11	Estee Lauder's	0.6028	5	1.08	Medium
9	the Ministry of Food and Drug Safety	0.6028	5	1.08	Medium
8	Bain Capital Private Equity	0.6028	5	1.08	Medium
4	CLIO	0.6028	5	1.08	Medium
7	Amorepacific's 2008	0.6028	5	1.08	Medium
6	Sephora, Target	0.6028	5	1.08	Medium



Risk Context: Common Topics

	topic_keywords	article_count
0	Outlier	719
6	bangladesh, run, wicket, mohammad, match	92
42	pcb, pakistan, cricket, team, coach	86
21	gladiator, peshawar, zalmi, king, quetta	68
37	oil, barrel, crude, price, cent	61
17	export, million, cotton, compared, year	57
35	murray, djokovic, wimbledon, federer, slam	50
39	oil, crude, price, singapore, cent	50
20	france, goal, germany, minute, euro	50
10	boxing, ali, fight, boxer, muhammad	48

- Organizations repeatedly exposed to anomalous coverage is high risk brands
- **Nature of risk:** Whether it's driven by negative sentiment, location mismatches, or spikes in volume.
- **Narrative context:** Which topics are linked to brand risk (e.g., energy, politics, health).
- **Actionable insight:** Stakeholders can see not just *that* they're at risk, but *why*.