

News-to-Narrative Pipeline Architecture (2025)

Data Sources

MarketAux API
5000+ sources

Alpha Vantage
Market Data

RSS Feeds
Real-time

Ingestion Layer

Rate Limiting

Deduplication

Date Extraction

Preprocessing

Text Cleaning

Language Detection

Entity Recognition

Embeddings

Sentence-BERT
384 dims

all-MiniLM-L6-v2

Normalization

Topic Modeling

UMAP
Reduction

HDBSCAN
Clustering

c-TF-IDF
Representation

Classification

FinBERT
Sentiment

Zero-shot
GPT-4o

Few-shot
Prompting

Aggregation

Rolling Windows

Exp. Smoothing

Outlier Detection

Output

Narrative
Time Series

Intensity
Scores

API
Endpoints