Real-Time Industry Insight System

This project provides a unified system that — for any publicly traded company — fetches company metadata, historical and real-time market data and related news; performs sentiment analysis on news flow; forecasts near-term stock price; detects unusual sentiment/price movements; and displays everything in a consolidated, user-friendly dashboard.

Main features:
- Fetch company details (description, founding date, industry) from Wikipedia.

- Collect historical and real-time market data via yfinance.

- Collect latest company/market news via RSS feeds.

- Perform sentiment analysis on news using a pre-trained model (FinBERT) and optionally leverage a large-language-model API (Gemini API) for additional analysis.

- Visualize sentiment distribution and market closing price over time.

- Use a financial-data API (Alpha Vantage) to gauge recent trends and flag bearish/ bullish signals.

- Forecast 7-day future price (or other metric) using time-series modeling (Prophet or ARIMA).

- Trigger alert via Slack when there's a sudden spike in sentiment or price (positive or negative).

- Provide a unified, intuitive dashboard combining all of the above.

Goal:

Provide investors/analysts/workflow-users with quick, automated insights combining fundamentals (company metadata), price history & forecast, and sentiment from news — all in one place — reducing manual effort and enabling data-driven decisions.

Data Sources & Data Handling

Company Metadata

The system retrieves company metadata by querying publicly accessible Wikipedia pages. A structured extraction process parses the page's info box and introductory sections to capture essential corporate details. The fields typically collected include:

- Company name and ticker symbol

- Sector and industry classification

- Founding year and key historical milestones

- Summary description of the company's business operations

This information is obtained programmatically using a combination of web-scraping utilities and Wikipedia APIs. Extracted data is standardized into JSON format and stored in the system's data directory for downstream use in the dashboard and analysis modules. The metadata collection ensures consistent formatting regardless of company page structure, integration with other modules.

Market Data

Market data ingestion is performed using the yfinance library, a robust interface for downloading historical and real-time financial datasets for publicly traded companies. The system retrieves:

- Daily OHLC (Open, High, Low, Close) prices

- Trading volume

- Adjusted close values

- Timestamps for all observations

Approximately 506 days of historical data are downloaded for each company to support trend analysis and time-series forecasting.

News Data / RSS Feeds

Recent company-related news is collected through RSS feeds sourced from reputable financial publishers and aggregators. Typical sources include industry news outlets, finance-focused portals, and global business publications. The system polls these feeds periodically and extracts key metadata such as:

- Article title

- Source

- Article link

- Short description or summary

Each news item is stored in structured JSON format. The consistent polling mechanism ensures the system always analyzes the most current information, supporting timely sentiment scoring and alert generation.

Sentiment Analysis

The sentiment analysis component processes each collected news article using a dual-model approach:

1. FinBERT Model:
A financial-domain-tuned transformer model that classifies text into Positive, Neutral, or

Negative sentiment. This ensures domain-appropriate interpretations of terminology commonly used in market-related news.

2. Gemini API (Optional Enhanced Analysis):
The system optionally enriches FinBERT outputs by performing deeper contextual sentiment assessment using Gemini. This includes summarization, confidence scoring, and refinement of sentiment labels where ambiguity exists.

The sentiment analysis pipeline includes several preprocessing steps:

- Removal of HTML tags and formatting artifacts

- Normalization of unicode characters

- Tokenization and text cleaning

Outputs include both categorical sentiment labels and numerical confidence scores. These are used in downstream trend analytics, visualization, and alerting. The system acknowledges inherent limitations such as model bias, interpretational ambiguity in short summaries, and language support constraints.

Time-Series Forecasting

The forecasting module predicts short-term stock price movements using two alternative time-series models:

1. Prophet:
A decomposable model suitable for capturing seasonality, trends, and holiday effects in financial time-series. Prophet is highly interpretable and resilient to missing values.

2. ARIMA:
A classical statistical model used when autocorrelation is significant and consistent patterns appear in residuals.

Workflow Summary:

1. Load 506 days of historical closing prices.

2. Handle missing values via forward-filling or interpolation.

3. Resample data to ensure daily consistency.

4. Fit the selected model to the historical series.

5. Generate a 7-day forecast with prediction intervals.

6. Export the forecast for visualization on the dashboard.

The forecasting engine provides both numerical outputs and plots for interpretability and strategic decision-making.

Trend / Technical Analysis via Alpha Vantage

To detect short-term bullish or bearish trends, the system incorporates technical indicators retrieved via the Alpha Vantage API. Used indicator:

- Simple and Exponential Moving Averages (SMA/EMA)

These signals are merged with sentiment scores to support more nuanced alerts and dashboard displays.

Alerting / Notifications

The alerting subsystem is designed to notify users when significant sentiment or market abnormalities occur. Alerts are triggered under conditions such as:

- Sudden spike in negative or positive sentiment across multiple news articles

- Sharp deviation in closing price compared to prior historical variability

- Detected bearish or bullish reversal signals from technical indicators

When an alert is triggered, the system formats a detailed message including:

- Company name

- Reason for alert (e.g., "High Negative Sentiment Spike Detected")

- Timestamp

- Relevant metrics (sentiment intensity, price deviation, etc.)

- Quick-access link to further analysis or dashboard view

Messages are delivered to a predefined Slack channel using a webhook URL configured via the system's .env file. This enables real-time team communication and rapid awareness of emerging market conditions.

Dashboard / Visualization

The system features a unified, user-friendly dashboard that synthesizes all upstream data. The dashboard is built to support interactive exploration of real-time and historical insights. Core components include:

Company Overview Panel

Displays company metadata such as name, sector, industry, headquarters, business description, and historical background. This gives users immediate context for the stock under analysis.

Market Data & Forecast Visualization

Shows historical closing prices over the 506-day window along with overlays such as:

- Moving averages
- Trend lines

- 7-day forecast generated via Prophet or ARIMA
- Confidence intervals for projected movement

This section supports visual interpretation of both past performance and expected short-term outcomes.

Sentiment Analytics Panel

Includes multiple visualization formats:

- Distribution of sentiment labels (Positive / Neutral / Negative)
- Sentiment intensity over time
- Sentiment-volume correlation charts (frequency of news vs sentiment strength)

Users can identify shifts in sentiment and correlate them with stock price variability.

Combined Price–Sentiment View

A synchronized dual-axis plot illustrates how sentiment trends align with price movement. This reveals patterns such as sentiment-driven volatility or price reactions to major news events.

Alerts Panel

Displays recently triggered alerts with:

- Timestamp
- Alert type
- Summary of reason
- Key metrics
- Direct navigation to related charts/data

This serves as the monitoring hub for rapid decision-making.