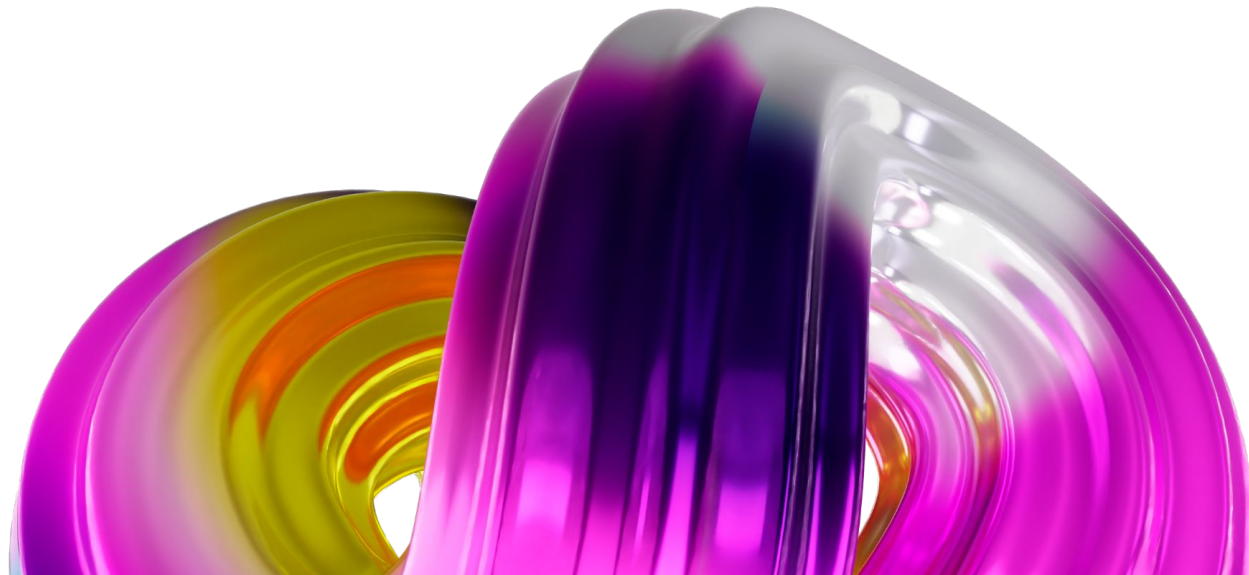


Can News Sentiment Predict Earnings Surprises?

Using sentiment analysis to uncover whether contrarian news articles can predict earnings surprises more accurately than the consensus.



01

Introduction

Avikalp Karrahe

UC Davis MSBA | Advanced Project Presentation | May 30, 2025





Problem Motivation

Earnings season brings high market volatility and investor anxiety. Financial news is flooded with articles echoing analyst consensus making it difficult to separate real insight from noise.

But sometimes, dissenting voices '**contrarian views**' get it right when the majority doesn't.

Our project asks: Can these contrarian signals, detected through sentiment analysis, help predict earnings surprises better than mainstream coverage?



02

Proposed Solution





Earnings season volatility

Earnings season presents significant fluctuations in financial markets due to the release of earnings reports. Investors and analysts closely monitor news to gauge sentiment, as this influences trading decisions.



Mainstream vs contrarian views

The majority of news articles often reflect a consensus viewpoint, which can lead to missed opportunities. Our study emphasizes the importance of identifying contrarian opinions, as these dissenting perspectives may offer more accurate predictions of earnings surprises.



Predicting earnings surprises

We explore how traditional forecasting techniques often fall short and highlight the potential advantages of integrating contrarian perspectives into predictive models.



03

Preview of Results



Extracting Earnings From Yahoo Finance



Earnings Calendar					
Find earnings for symbols					
Showing Earnings for: AAPL					
Clear					
Symbol	Company	Earnings Date	EPS Estimate	Reported EPS	Surprise (%)
AAPL	Apple Inc.	April 29, 2026 at 6 AM EDT	-	-	-
AAPL	Apple Inc.	January 28, 2026 at 4 PM EST	-	-	-
AAPL	Apple Inc.	October 29, 2025 at 6 AM EDT	-	-	-
AAPL	Apple Inc.	July 30, 2025 at 6 AM EDT	-	-	-
AAPL	Apple Inc.	May 1, 2025 at 4 PM EDT	1.63	1.65	+1.41
AAPL	Apple Inc.	January 30, 2025 at 4 PM EST	2.35	2.4	+2.15
AAPL	Apple Inc.	October 31, 2024 at 4 PM EDT	1.6	1.64	+2.35
AAPL	Apple Inc.	August 1, 2024 at 4 PM EDT	1.35	1.4	+3.99
AAPL	Apple Inc.	May 2, 2024 at 4 PM EDT	1.5	1.53	+1.97

Extracting Apple Articles Through Guardian API



A	B	C	D	E	F
Body	Headline	Subheadline	Date	Short URL	Author
Nvidia beat Wa	Nvidia beats Wall Street ex	Chip-manufacturing compa	2025-05-28 19:54:03	https://w	Johana Bhuiya
The boss of the	Nvidia will spend hundreds	Chipmaker's promise is a si	2025-03-20 09:59:17	https://w	Kalyeena Mak
This is the lates	The Thinking Machine: Jens	Stephen Witt's entertaining	2025-04-20 14:00:09	https://w	John Naughton
US chip exports	US chip export controls are	Comments from Jensen Hu	2025-05-21 05:25:16	https://w	Helen Davidso
The US Federal	Fed chair says Trump tariffs	Value of Nvidia dropped by	2025-04-16 15:21:37	https://w	Lauren Almeida
The chief execu	Nvidia's CEO makes surpris	Jensen Huang causes stir o	2025-04-17 11:25:04	https://w	Helen Davidso
US stock marke	US stock markets see worst	All three major US index fu	2025-04-03 15:32:54	https://w	Lauren Aratan
US stock marke	US stock markets expected	Exemption, seen as a climb	2025-04-13 17:36:35	https://w	Phillip Inman
Hello, and welc	From spit to bankruptcy: th	Plus: Nvidia bets on AI-pow	2025-03-25 13:17:01	https://w	Blake Montgo
On his tour of t	How Donald Trump's 'histo	The deals stand to enrich t	2025-05-17 12:00:06	https://w	Blake Montgo
Morning everyo	Morning Mail: Trump tariffs	Traders rush to sell stocks	2025-04-03 19:46:09	https://w	Martin Farrer
China has accus	China accuses UK politician	Embassy criticises 'slander	2025-04-16 12:53:56	https://w	Jasper Jolly
Growing uncert	More than £2bn wiped from	Shares in Bunzl, which prov	2025-04-16 10:17:01	https://w	Lauren Almeida
Hello, and welc	Trump the AI salesman mak	The president sold the US r	2025-05-20 14:01:55	https://w	Blake Montgo

Sentiment Analysis of Walmart Articles Using Meta-Llama3

A	B	C	D	E	F	G	H	I	
Body	Headline	Subheadline	Date	Short URL	Authors	URL	article_words	Sentiment	
Walmart on Th	Walmart shares fall upon disappointing forecast, fueling	World's largest retail	2025-02-20 16:16:07	https://www.theguardian.com/p/xxtpkb	Guardian s	https://wv	394	Negative	
Walmart will ha	Walmart will have to raise prices this month due to Tru	Retailer expects ann	2025-05-15 12:44:50	https://www.theguardian.com/p/x29fe7	Lauren Ara	https://wv	474	Negative	
It's not funny, c	The truth is finally dawning on Britain: toadying to Tru	Jolly humouring and	2025-05-01 07:00:18	https://www.theguardian.com/p/x26bq5	Emma Bro	https://wv	814	N/A	
Last month Dor	What is the DEI rollback and where will it spread next? Many large US corpo		2025-02-14 15:04:34	https://www.theguardian.com/p/xxhdnh	Joanna Pai	https://wv	986	Negative	
Asda is to inves	Asda to invest in price cuts to battle drop in sales and i	Supermarket chain's	2025-03-14 14:00:24	https://www.theguardian.com/p/xxn7zk	Sarah Butl	https://wv	709	Negative	
As economic ur	Gold has surged amid economic uncertainty. Should yc	As Trump escalates a	2025-03-22 11:00:49	https://www.theguardian.com/p/xxzjmh	Adam Gab	https://wv	776	Negative	
Donald Trump a	Trump's tariffs will be paid by the poor – while his tax	The math doesn't wo	2025-03-15 11:00:48	https://www.theguardian.com/p/xxn97e	Robert Rei	https://wv	1332	Negative	
Beth Benike kn	US small business owner says China tariffs endanger h	Beth Benike, whose i	2025-04-10 10:00:27	https://www.theguardian.com/p/x2233g	Rachel Lei	https://wv	1334	Negative	
One in five Ame	Exclusive: 20% of Americans support boycott of firms : New poll also shows		2025-03-22 10:00:47	https://www.theguardian.com/p/xxzeqk	Lauren Ara	https://wv	683	Negative	
Gina Rinehart h	How rich is Gina Rinehart, and how much will she earn	Australia's wealthies	2025-04-13 15:00:19	https://www.theguardian.com/p/x2xyvx	Luca Ittim	https://wv	934	N/A	
On Thursday m	Rev Al Sharpton meets with Target CEO over halting of	Activists had called f	2025-04-17 21:24:51	https://www.theguardian.com/p/x23yq5	Adria R Wa	https://wv	366	Negative	
In late January,	'They've lost my trust': consumers shun companies as	Americans are using	2025-02-28 11:00:08	https://www.theguardian.com/p/xxka39	Lauren Ara	https://wv	873	Negative	
British business	UK firms ditching diversity and inclusion 'face higher ris	Top employment law	2025-04-07 11:05:43	https://www.theguardian.com/p/x2xbe5	Sarah Butl	https://wv	440	N/A	
A rollback of di	Rollback on diversity policies 'risks undoing decades of As Donald Trump rev		2025-02-14 15:14:45	https://www.theguardian.com/p/xxhfqv	Chris Osuh	https://wv	916	Negative	
Exports to the C	Chinese e-commerce exports to US plummet by 65% in	Temu and Shein amo	2025-04-30 13:00:52	https://www.theguardian.com/p/x26bxa	Lisa O'Car	https://wv	545	Negative	
Donald Trump s	Trump extends deadline for TikTok sale to non-Chinese	Deadline set by US p	2025-04-04 18:15:07	https://www.theguardian.com/p/x2x2x2	Dara Kerr	https://wv	482	Positive	

04

Methodology



A detailed approach for market sentiment analysis and predictions

Step 1: Extract Key Dates

Gather earnings dates, forecasted EPS, and actual EPS from Yahoo Finance to establish a timeline for analysis.

Step 2: Collect News Articles

Retrieve articles from the Guardian API (filtered by the technology tag) and scraped sites like Gizmodo, Verge, and Wired while filtering out irrelevant content.

Step 3: Sentiment Analysis with Llama 3

Use a local open-source model (Llama 3 GGUF) to assign positive, negative, or neutral sentiment to each article, providing insights into market sentiment.



A detailed approach for market sentiment analysis and predictions

Step 4: Contrarian Detection

Determine the majority sentiment for each company. Articles that disagree with this dominant sentiment are flagged as contrarians, helping uncover potential market anomalies.

Step 5: Outcome Evaluation

Compare contrarian sentiment against EPS surprise direction (Actual vs. Forecast).

Optionally, validate further using short-term stock price movement to assess predictive power.



05

Data



Model Selection & Validation



To evaluate the accuracy of open-source models, we compared the **Meta Llama 3 model** with **OpenAI's GPT-4o-mini** using a sample of 30 articles.

The two models showed approximately **90% agreement** in sentiment classification, which gave us confidence to proceed with Llama 3.

We also manually reviewed several articles to ensure both models were producing reliable and consistent results.

Since the Llama 3 API had rate limits, we implemented `time.sleep()` in our code to prevent throttling during batch processing.

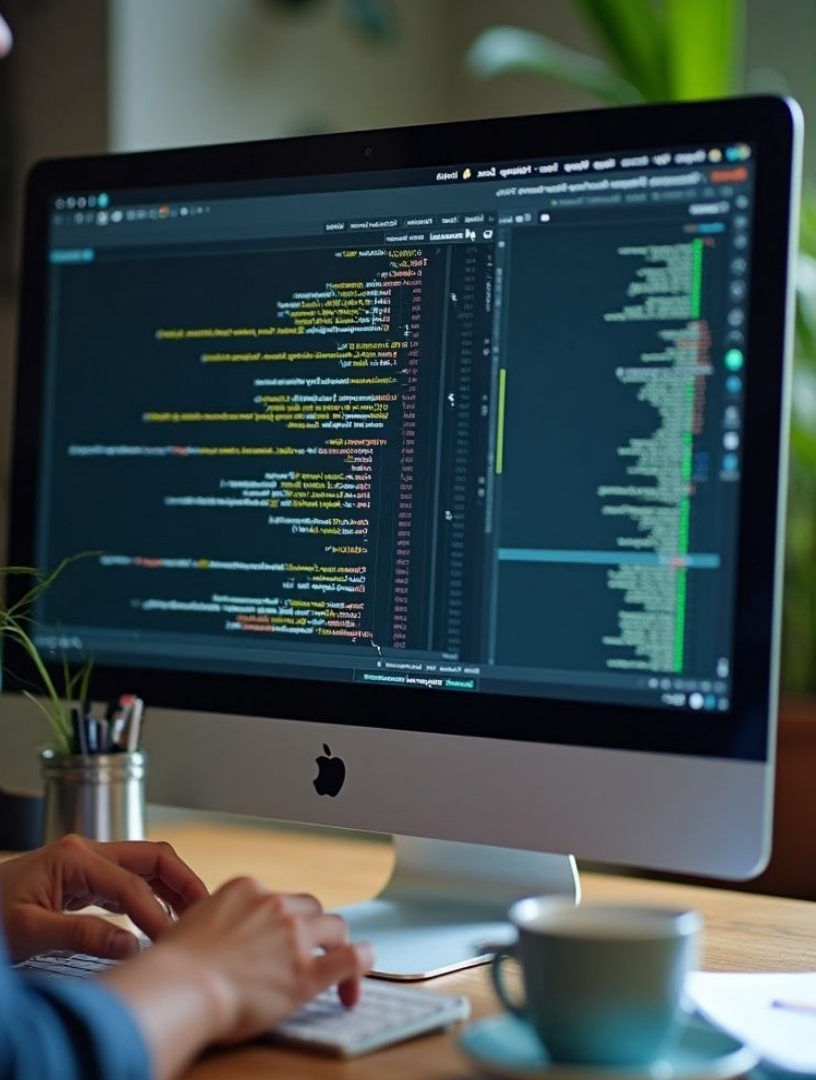
Problems with NewsAPI



NewsAPI didn't return full article content — only headlines and brief summaries.

The free tier also restricted access to historical data, making backtesting difficult.

We attempted web scraping as a workaround, but the structure of each site varied widely, making it inconsistent and time-consuming.



Custom scraping solutions

As a result, we pivoted to using the Guardian API and also scraped articles from Wired, Verge, and Gizmodo.

To overcome the limitations of existing APIs, we developed custom web scrapers to gather data from various reputable sources. This approach improved content diversity but introduced new challenges, such as inconsistent formatting and filtering irrelevant articles.



Guardian API Challenges & Workarounds



Challenge 1: Irrelevant Results for Company Names

Searching “Apple” returned results like apple pie recipes or unrelated people named Apple.

Solution: Filtered articles by the “technology” tag to narrow to relevant company news.

Challenge 2: Noisy Blog-Style Content

Many articles were live blogs or updates with little substance and no clear author.

Solution: Only included articles that had a definite author and excluded blogs from analysis.

Guardian API Challenges & Workarounds



Challenge 3: Limited Coverage

As a UK-based publication, the Guardian lacked coverage on some major U.S. tech firms.

Solution: We plan on using supplemented Guardian data from other free or paid publication firms.

Despite these limitations, we extracted a high-quality dataset by carefully filtering and augmenting Guardian results.

06

Results



Company-Level Results & Insights

Apple (often 0%)

Alignment was **consistently low**.

Media sentiment may lag behind or contradict actual earnings, offering limited predictive value.

ExxonMobil (~77%)

Only Q1 2025 showed **significant alignment**.

Energy sector sentiment may be influenced by environmental or geopolitical coverage bias.

NVIDIA (~85%)

Strong alignment during recent quarters, especially when EPS beats were high.

Tech media appears more responsive and aligned with market performance.



Company-Level Results & Insights

Pfizer (~60%)

Mixed results across quarters.

Healthcare coverage is likely influenced by clinical and regulatory cycles, which do not always align with earnings outcomes.

Walmart (~50-60%)

Moderate alignment throughout.

Retail sentiment has moderate predictive power but is often affected by broader economic signals.



07

Summary



Workflow Summary – All Companies

1. Data Structure

News articles and earnings data were processed quarterly. Each Excel sheet included articles tagged with sentiment (Positive, Negative, Neutral) alongside quarterly EPS data.

2. EPS Surprise Rule

- If actual EPS > forecast → expected sentiment is positive
- If actual EPS < forecast → expected sentiment is negative

3. Contrarian Detection Logic

- If sentiment matches EPS surprise → Aligned
- If sentiment differs from EPS surprise → Contrarian
- Alignment percentage was calculated for each quarter.





Final Takeaways

Most financial news mirrors consensus.
But a small subset of contrarian views may contain more accurate predictive signals.

Our pipeline detects those signals.
By comparing article sentiment with actual earnings results, we identify when dissenters got it right.

Contrarian accuracy varies by sector.
Tech companies like NVIDIA showed strong alignment.
Others, like Apple and ExxonMobil, did not.

Open-source models can deliver.
Llama 3 matched GPT-4o sentiment 90% of the time, enabling cost-effective analysis at scale.

Next step: Full automation.
We plan to deploy a workflow system that automates article ingestion, sentiment scoring, and signal detection before earnings calls.





Thank you!

Do you have any questions?