Data Engineering Project Proposal

뉴스데이터를 활용한 LLM 기반 포트폴리오 최적화



6조

박장환 | 윤원상 | 장미현 | 정유진 | 최인주

Contents

01 Problem Statement

02 Architecture

03 Data Collection

04 Data Preprocessing

05 Modeling

06 Result

01. Problem Statement

트럼프·머스크 파국에 테슬라 주가 14% 폭락...시총 1조 달러 붕괴

약 한 달 만에 주가 300달러 하회
"실망" vs "나 없었으면 선거 패배"

정수민 기자 | 입력 : 2025.06.06 14:18:09



일론 머스크 테슬라 CEO(오른쪽)가 미국 공화당 대선 후보 도널드 트럼프의 펜실베이니아주 유세 현장에 참석해 연설에 나서고 있다. (출처=UPI 연합뉴스)

도널드 트럼프 미국 대통령과 일론 머스크 테슬라 최고경영자(CEO) 간 갈등이 격화되면서 6월 5일(현지 시각) 테슬라 주가가 폭락했다.

이날 뉴욕증시에서 테슬라 주가는 전일 대비 14.26% 하락한 284.70달러(38만6309원)로 마감했다. 장중 한때 273.21달러까지 떨어지며 17% 이상 급락하기도 했다.

테슬라 주가는 지난 5월 9일(298.26달러) 이후 약 한 달 만에 300달러 아래로 내려갔다. 시가총액은 9170억달러로 줄며 1조달러 선이 무너졌다. 하루 만에 약 1520억달러(약 206조원)가 증발한 셈이다.

PROBLEM

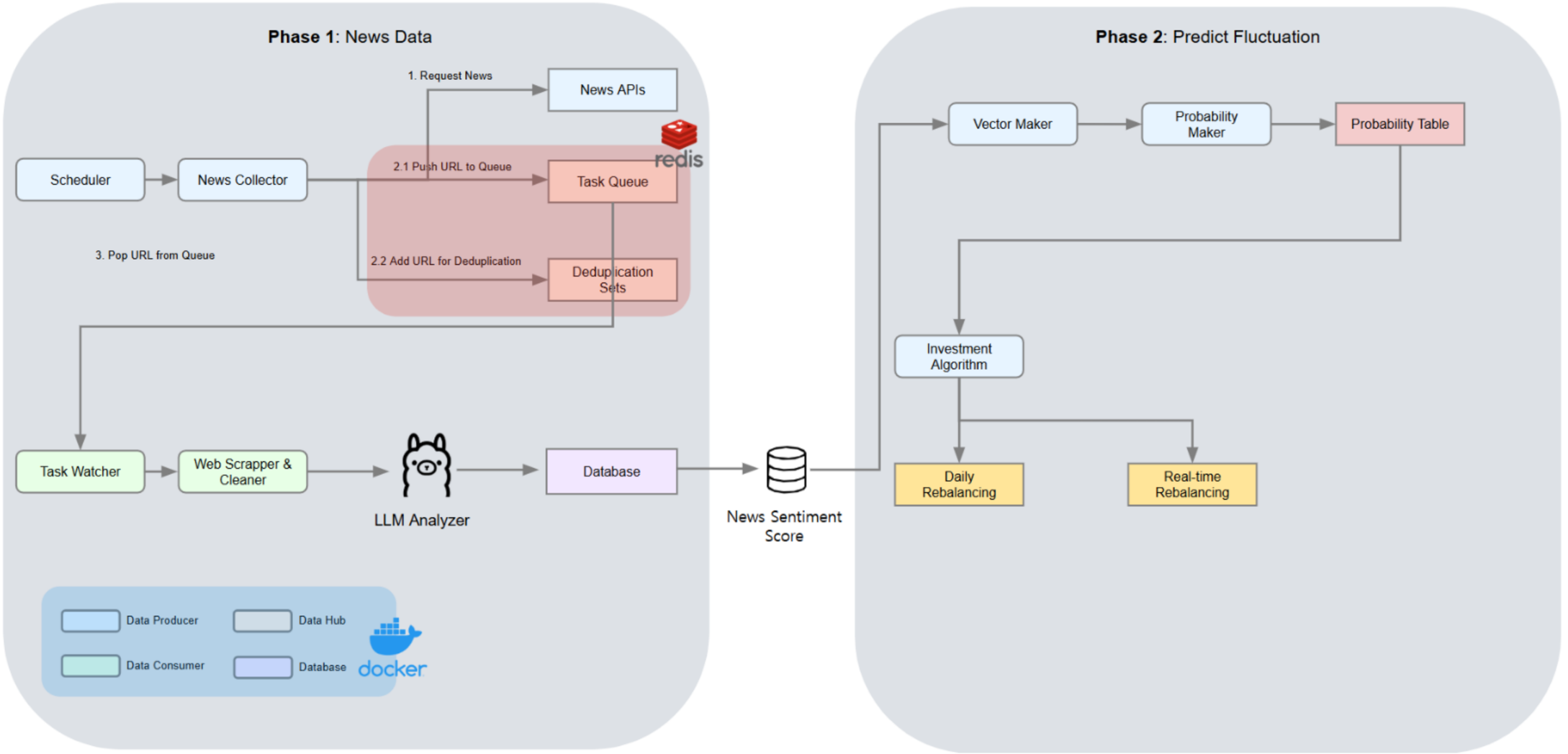
주가 영향력이 높은 비정형 데이터의
복잡성 및 방대함으로
개인투자자의 수집/분석/반영에 한계

SOLUTION

뉴스데이터 & LLM 활용

Nasdaq 100 대상 포트폴리오 최적화

02. Architecture



03. Data Collection

카테고리	활용데이터	구성	범위	출처
주가 데이터	과거 주가데이터	시가, 고가, 저가, 종가, 거래량	최근 5년, NASDAQ100	Yahoo Finance
	주가데이터 API	시가, 고가, 저가, 종가	실시간, NASDAQ100	polygon.io
뉴스 데이터	과거 뉴스데이터	기사 전문	최근 5년, 경제뉴스	GDELT
	뉴스데이터 API	기사 전문	실시간, 경제뉴스	polygon.io

04. Data Preprocessing

Step 1 '주가데이터-뉴스데이터' 시점 통일

Step 2 LLM을 활용하여 '뉴스데이터-종목&감성점수' 라벨링

INPUT

```
너는 금융 분석 전문가야.
...
  [뉴스 기사]:
  {How the US got left behind
   in the global electric car race ··· }
...
```

OUTPUT

```
{
  "ticker_symbol": "TSLA",
  "sentiment_classification": -1,
  "sentiment_score": -0.65,
  "relevance_score": 0.9
}
```

05. Modeling

Step 1 Vector Maker (Score to Vector)

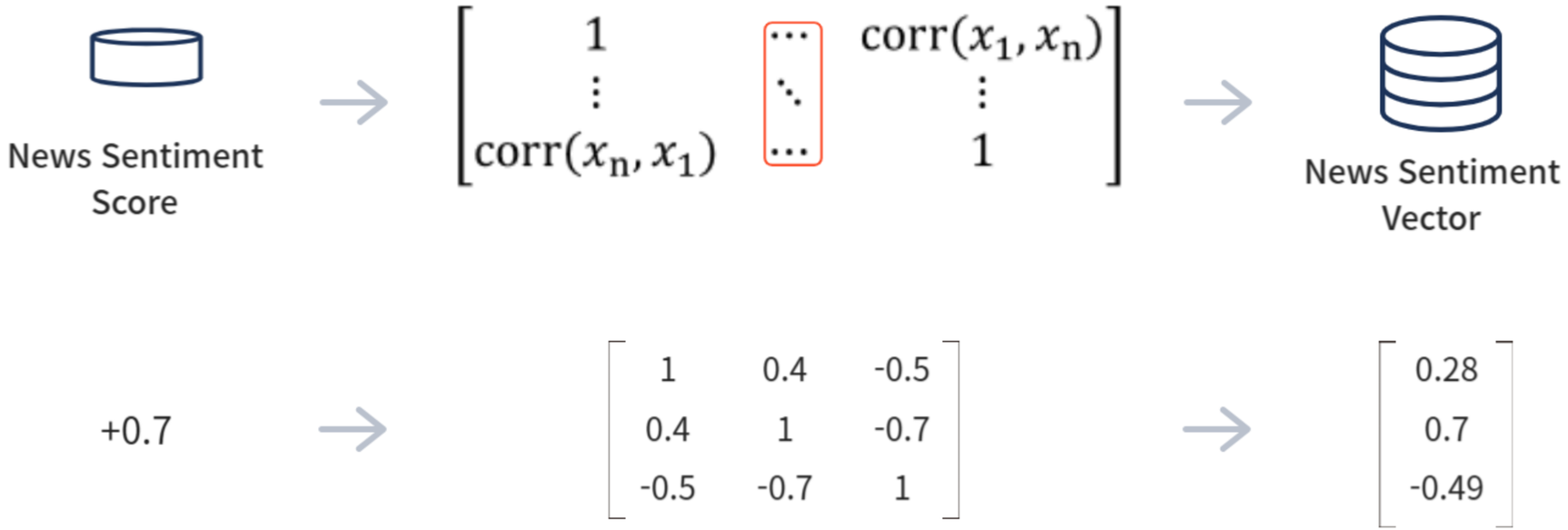


$$r_i = \alpha_i + \beta_i \cdot r_m + \epsilon_i$$

$$\begin{bmatrix} \mathrm{corr}(x_1, x_1) & \cdots & \mathrm{corr}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathrm{corr}(x_n, x_1) & \cdots & \mathrm{corr}(x_n, x_n) \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \cdots & \mathrm{corr}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathrm{corr}(x_n, x_1) & \cdots & 1 \end{bmatrix}$$

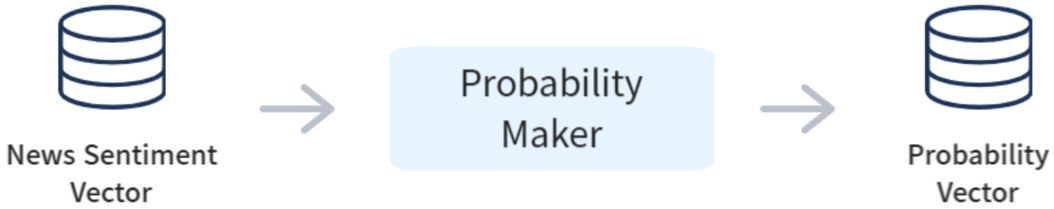
05. Modeling

Step 1 Vector Maker (Score to Vector)



05. Modeling

Step 2 Probability Maker (Vector to Probability)

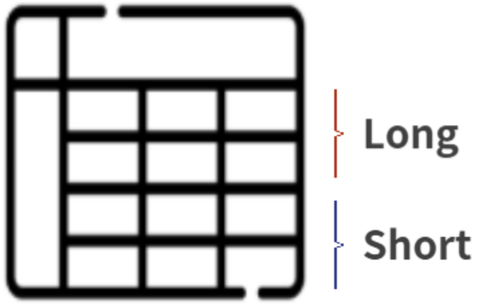


Step 3 Probability Table



06. Result

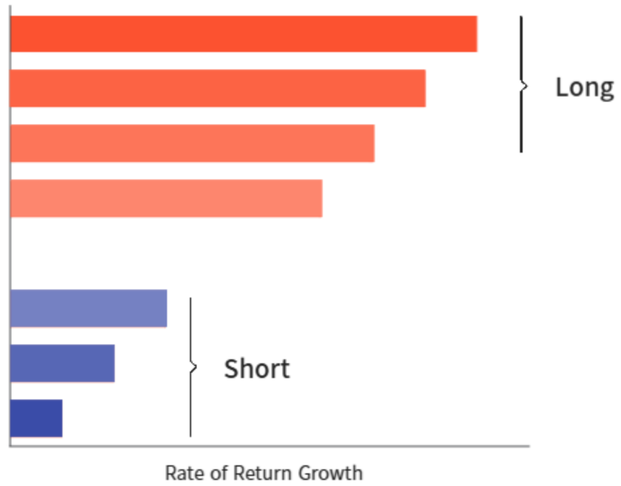
실시간으로 예측된 주가 방향성을 기반하여 **상승률 기준 Nasdaq100 종목 랭킹**
→ **포트폴리오 최적화** 전략 제시



Probability Table



큰 충격으로 예상되는 뉴스 발표 시
실시간 리밸런싱



Data Engineering Project Proposal

THANK YOU

6조

박장환 | 윤원상 | 장미현 | 정유진 | 최인주