

Dolování dat z webových stránek s využitím AI nástrojů

Semestrální projekt

Bc. Ondřej Kod'ousek

Vysoká Škola Baňská - Technická Univerzita Ostrava
Fakulta Elektrotechniky a Informatiky

27. května 2025

Vedoucí práce: Doc. Ing. Gajdoš Petr, Ph.D.

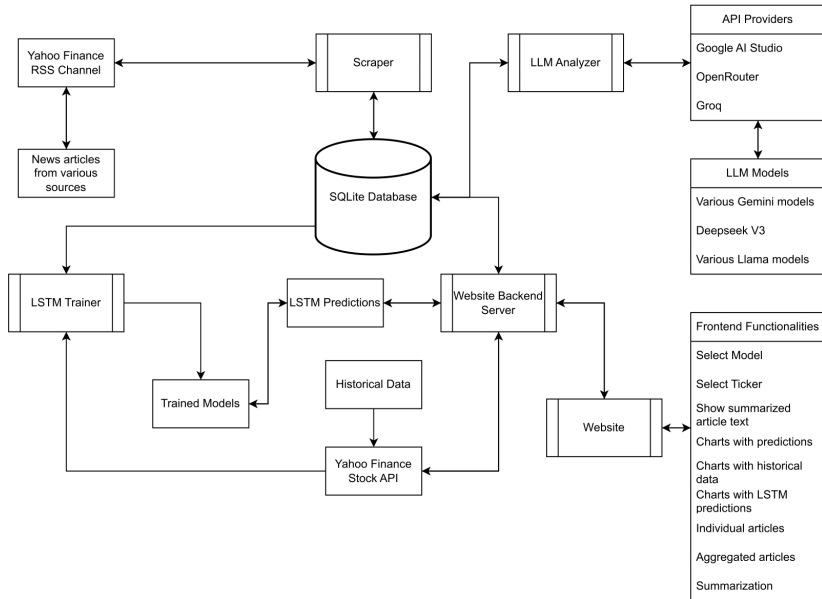
Cíl projektu

- Analýza sentimentu textu pro predikci změn cen akcií pomocí LLM
- Ověření praktické použitelnosti tohoto přístupu
- Vyhodnocení úspěšnosti predikcí

Postup

- 1 Automatizovaný sběr dat z finančních zpravodajských serverů
- 2 Analýza textu pomocí různých LLM modelů
- 3 Porovnání s LSTM predikcemi a historickými daty
- 4 Vizualizace výsledků ve webové aplikaci

Architektura projektu



Ukázka webové aplikace

gemin-2.5-flash-preview-04-17
gemin-2.5-pro-preview-03-25
gemin-2.0-flash
gemin-2.0-flash-lite
gemin-1.5-flash
gemin-1.5-flash-8b
gemin-1.5-pro
deepseek/deepseek-chat-free
llama-3.3-70b-versatile
llama-3.1-8b-instant
gemma2-9b-it
distil-whisper-large-v3-en
deepseek-r1-distill-llama-70b
meta-llama/llama-4-maverick-17b
meta-llama/llama-4-scout-17b-16e
qwen-qwq-32b

Minimum articles: 3

Individual articles

Aggregated graph

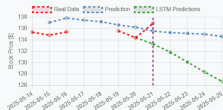
Show metrics

Batch download

LOID (Lucid Motors)
LLY (Eli Lilly)
LYFT (Lyft)
MAT (Mattel Inc.)
META (Meta Platforms, Inc.)
MO (Altria Group)
MP (MP Materials Corp)
MRVL (Marvell Technology, Inc.)
MSCI (MSCI Asia-Pacific Index)
MSFT (Microsoft Corporation)
MSTR (MicroStrategy Incorporated)
NA (Mortgage rates)
NBIS (Nebius Group N.V.)
NCLH (Norwegian Cruise Line Holdings Ltd.)
NFLX (Netflix)
NRG (NRG Energy)
NU (Nu Holdings)
NVDA (Nvidia Corp)
NVO (Novo Nordisk)
O (Realty Income Corporation)
ORCL (Oracle Corporation)
ORLY (O'Reilly Automotive)
PANW (Palo Alto Networks)
PEP (PepsiCo, Inc.)
PINS (Pinterest, Inc. (NYSE:PINS))
PLTR (Palantir Technologies Inc.)
PLUG (Plug Power Inc.)
PR (Permian Resources Corporation)
PSX (Phillips 66)
QCOM (QUALCOMM Incorporated)
RBLX (Roblox)
RDDT (Reddit, Inc.)
REGN (Regeneron Pharmaceuticals, Inc.)
RIVN (Rivian Automotive, Inc.)
RTX (RTX Corporation)
SBUX (Starbucks Corp)
SHOP (Shopify)
SMCI (Super Micro Computer)
SOFI (Sofi Technologies, Inc.)
SONY (Sony Group Corporation)
SOUN (SoundHound AI)

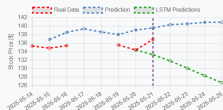
[Yahoo Finance: Nvidia stock extends gains as Saudi Arabia set to spend billions on AI chips, US moves to rescind Biden's chip curbs](#)

The article discusses Nvidia's recent stock performance and positive developments, including a significant deal with Saudi Arabia and potential easing of export restrictions. Analyst upgrades and positive market sentiment are also highlighted, contributing to the bullish outlook for NVDA. The author projects a positive trend for the stock, considering these factors.



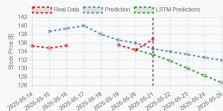
[Insider Monkey: NVIDIA Corporation \(NVDA\) Secures Major AI Chip Deal with Saudi Arabias HUMAIN](#)

The article discusses NVDA's position in the AI sector, especially in light of the changing US policy towards AI chip exports to the Middle East, specifically Saudi Arabia. The company is expected to sell a significant number of AI chips, including its latest Blackwell chips, to the region. The article suggests a positive outlook but doesn't provide explicit price predictions. The information from the article allows me to make positive predictions, as the change in US policy may have impact on stock growth.



[MarketWatch: Wall Streets biggest bull held his nerve throughout this years selloff. What hes saying now.](#)

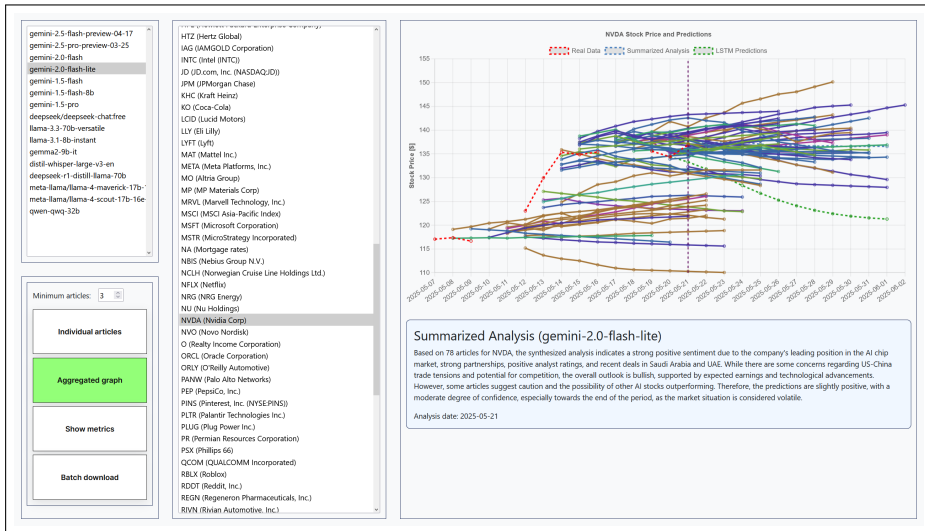
The article discusses Nvidia's strong performance, referencing its past successes and a recent partnership with a Saudi Arabian AI firm. It highlights the potential of AI and suggests a positive outlook for Nvidia due to the evolving regulatory environment and expected M&A activity. I expect positive growth for Nvidia in the coming days, with some correction in the final days of prediction.



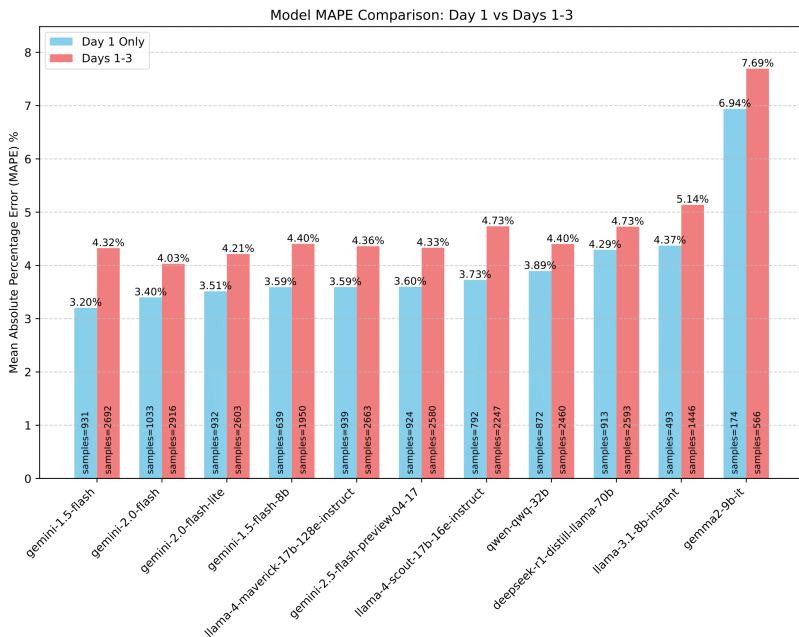
[Reuters: US close to letting UAE import millions of Nvidia's AI chips.](#)



Ukázka webové aplikace

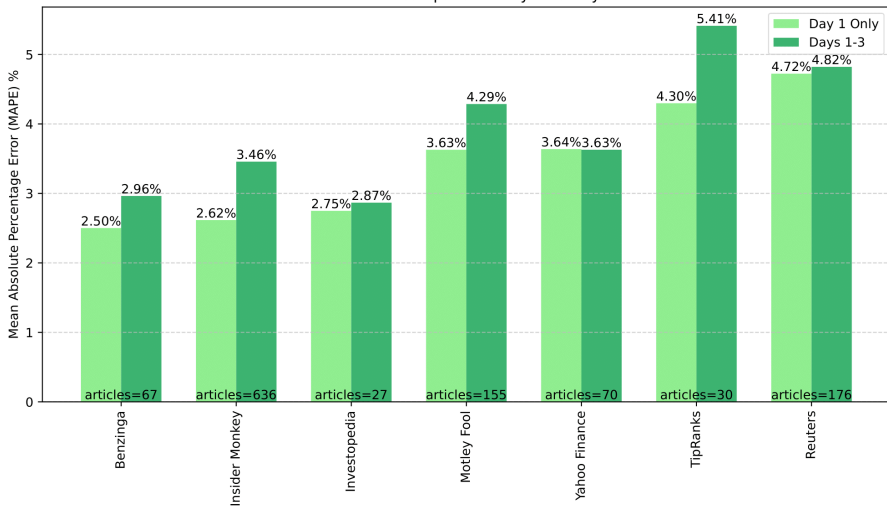


Výsledky modelů

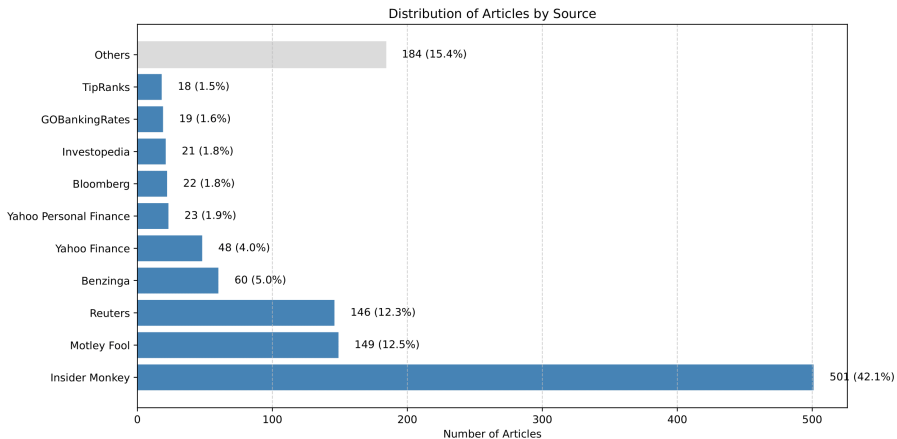


Výsledky zdrojů

Source MAPE Comparison: Day 1 vs Days 1-3



Poměr zdrojů dat



Hlavní zjištění

- Současné modely lze do jisté míry využít pro predikci pohybu na burze
- Rozdíly mezi modely jsou menší než očekávané
- Zdroje článků mají výraznější vliv na kvalitu predikcí než samotné modely
- Potencionálně lepší výsledky než LSTM (5.2% proti 3.2% MAPE v první den)

Možná vylepšení

- Využití specializovaných modelů (např. FinBERT)
- Kombinace s dalšími predikčními technikami
- Více vyrovnaný poměr zdrojů článků
- Zaměření na konkrétní zdroje článků a modely pro co nejlepší výsledky

Děkuji za pozornost

Bc. Ondřej Kod'ousek
VŠB-TUO, Fakulta Elektrotechniky a Informatiky
27. května 2025