Meme Coin Analysis Dashboard

Data Engineering Project Proposal



much data.
such pipeline.

IDSC501

Team12 members: Jaeho Kim, SeungJun Oh, Inha Jung

Doge illustration: Icons8

Contents

01 **Ideation** — Why predict meme coin prices?

02 **Data collection** — Crawling, APIs and source coverage

03 **Pipeline & preprocessing** — Frozen inputs for the model

04 **Database & dashboard** — Predictions, evidence and results

Why meme coins?

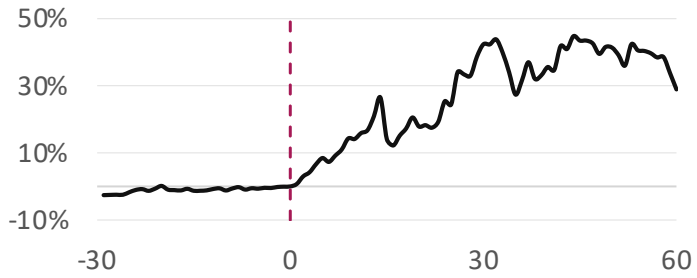
When they post

X / 04 Feb 2021



“Doge”

DOGE / USDT **+28.9% in 60 min**



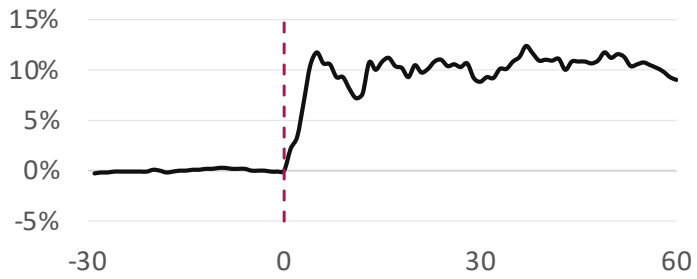
Capture: Dablander / Know Your Meme

Truth Social / 23 Mar 2025



“I LOVE $TRUMP”

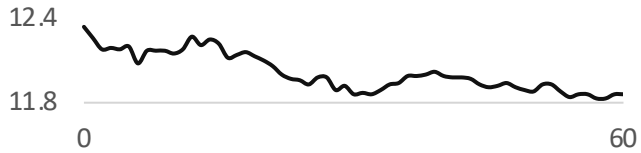
TRUMP / USDT **+9.0% in 60 min**



When we buy…



TRUMP after the spike



16:09–17:09 UTC, 23 Mar 2025. USDT.
Meme caption, not our trade.

Can prices + posts predict the next move?

Binance 1m; X: minutes. Selected events, not causal proof. Captures: KYM / The Block. Art: UniqueDesignOK.

Data collection

14,773 price rows

10 coins · social collection planned

Source	Collection method	Output
CoinMarketCap Historical Data	**Web crawling**	Daily price CSV
X via FxTwitter	**Third-party API**	Posts + reactions in JSON
Reddit via Arctic Shift	**Archive API / files**	Posts + comments

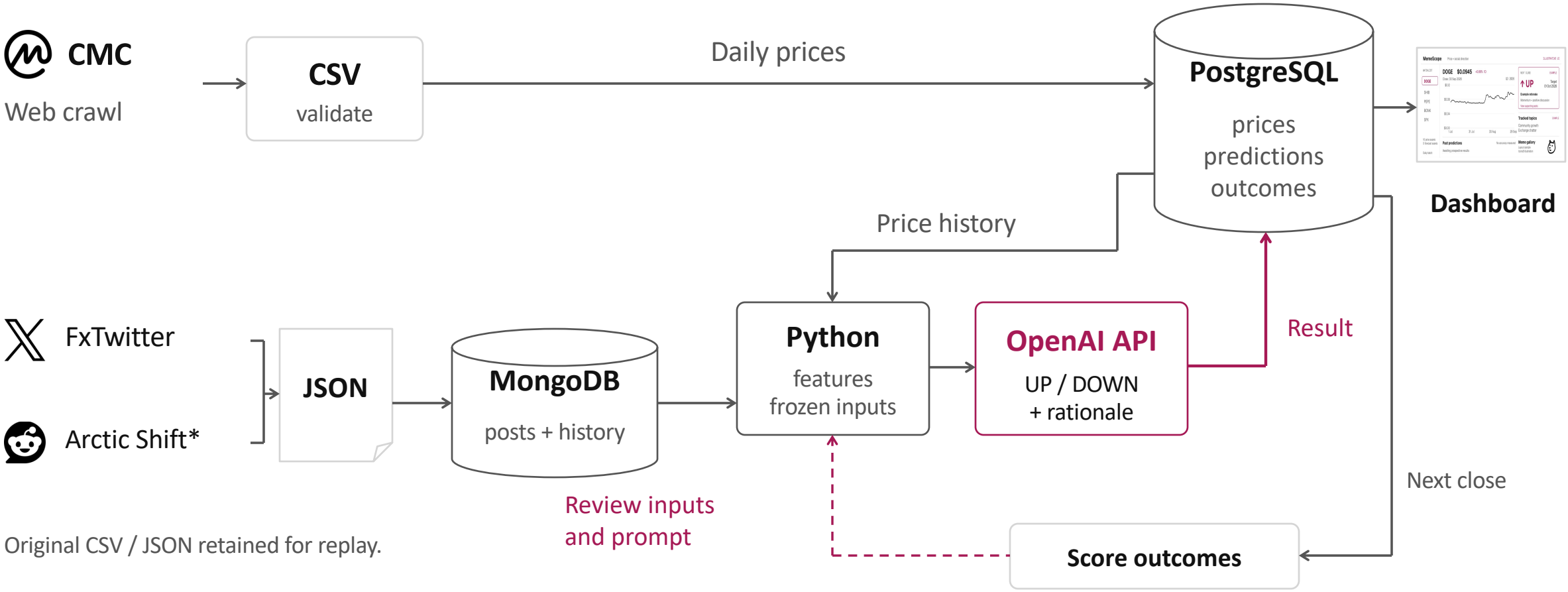
Daily social batches, with a seven-day overlap.

Project websites: crawl links and contract details to verify sources.

Reddit awaits access conditions. Price refresh needs new CSVs or restoration of the team crawler.

Data pipeline

Next UTC close: UP or DOWN



Original CSV / JSON retained for replay.

Dashed: team review for future versions. Post details read MongoDB. * Reddit: conditional access.

Data preprocessing

Only information available when the forecast is issued

Price features

31 daily closes
Returns and volatility
Volume changes

Social features

Activity by source
Sentiment + coverage
Selected post excerpts

Frozen snapshot

Known before cutoff
Stable IDs and hashes
Model + prompt version

UTC normalization · ID deduplication · verified coin matching

No future information

Later posts and updated reactions stay out of earlier inputs.

Freeze at 03:00 UTC. Publish by 03:15. Missing data or failed inference remains unavailable.

Why two databases?

PostgreSQL can also store JSON.

MongoDB

Social documents

Posts and observations
Platform-specific fields
Model input / output

PostgreSQL

Relational analytics

Prices and daily features
Predictions and outcomes
Keys, constraints and joins

Document model → ETL → relational model

Document storage and relational analysis, connected by a traceable ETL.

Trade-off: two databases, indexes, backups and consistency checks.

Prediction dashboard

MemeScope Price + social direction ILLUSTRATIVE UI

WATCHLIST

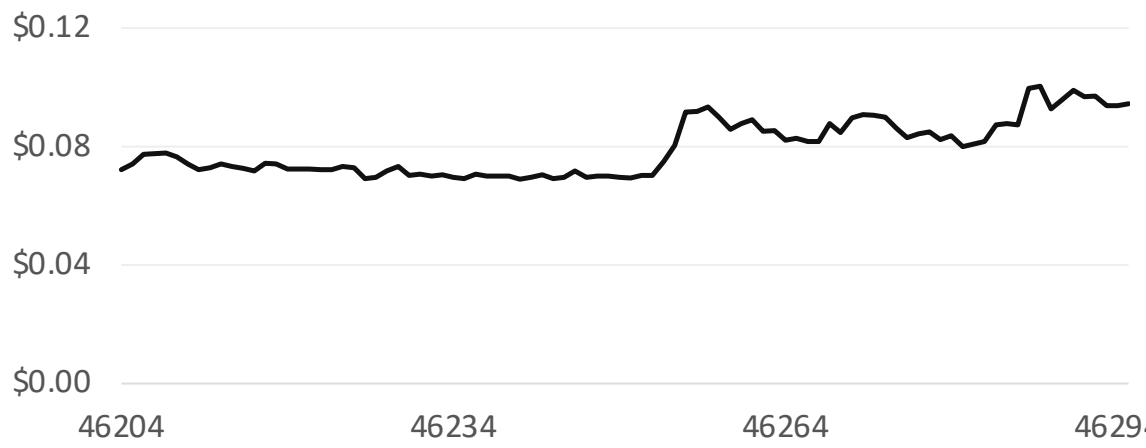
- **DOGE**
- SHIB
- PEPE
- BONK
- SPX

10 price assets
5 forecast assets

Daily batch

DOGE $0.0945 +0.66% 1D

Close: 30 Sep 2026 Q3 2026



Past predictions

No accuracy measured

Awaiting prospective results

NEXT CLOSE EXAMPLE

↑ UP

Target
01 Oct 2026

Example rationale

Momentum + positive discussion

View supporting posts

Tracked topics

EXAMPLE

Community growth
Exchange chatter

Meme gallery

Layout sample
Icons8 illustration



Supplied DOGE prices. Prediction and social text are examples. Image: Icons8.