ZENPULSAR
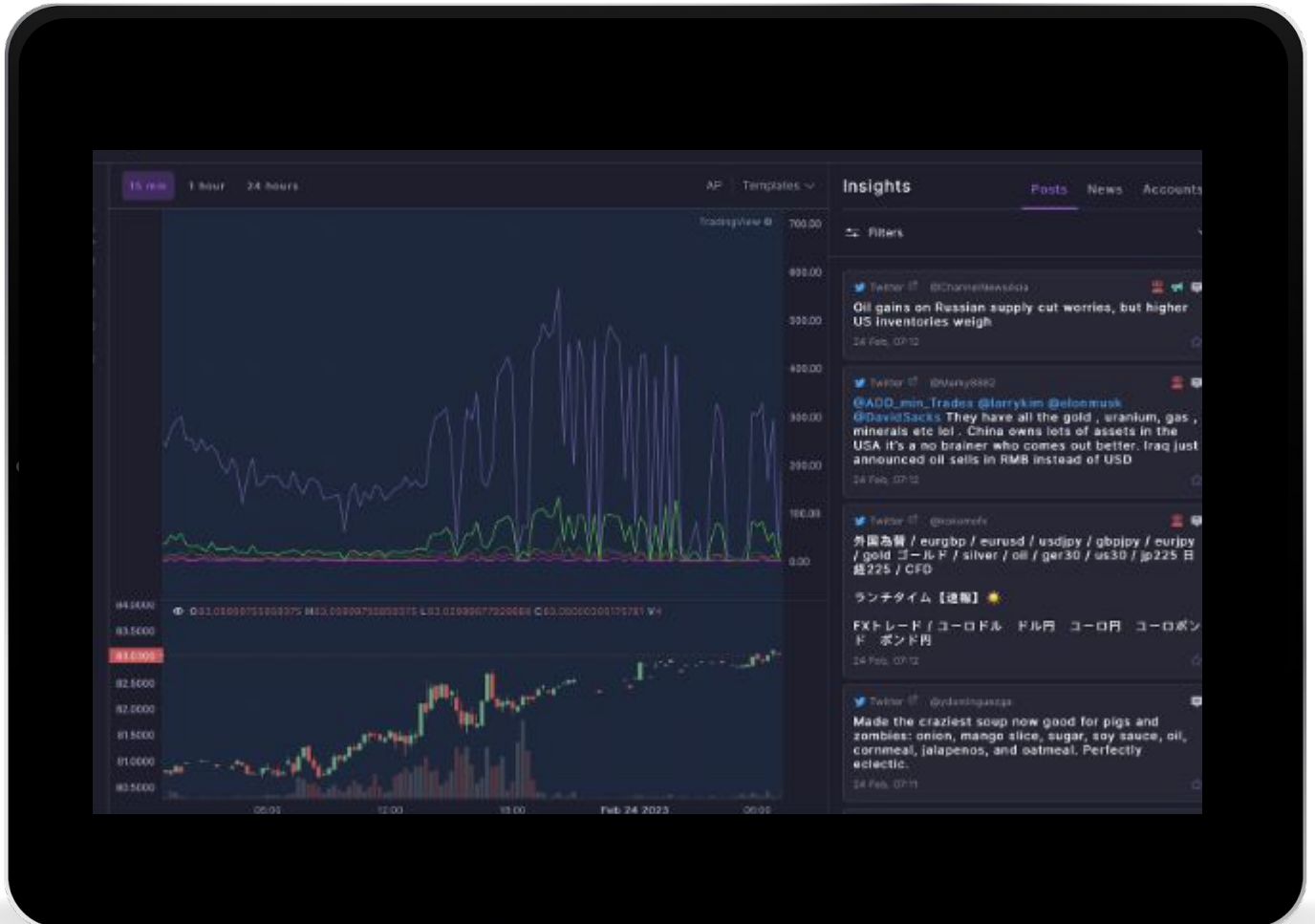
Social Media Pulse Data Set
Commodities

ZENPULSAR

ZENPULSAR's PUMP Social Media Pulse for Commodities tracks and quantifies the impact of social media on commodities. This unique data set generates ALPHA, providing a detailed analysis of the activities of influencers, financial professionals, retail investors, and bots across Social Media platforms.

ZENPULSAR's data centric AI platform — "PUMP" — monitors in real time multiple social media networks to track activities related to financial and crypto assets and then to analyse them. It detects emerging viral narratives likely to form trends and impact financial assets. PUMP clears out the noise of social media with unmatched speed and accuracy. It identifies viral narratives related to the assets you track, early signals we can spot and act on before the crowds and everyone else.

ZENPULSAR's technology is also leverage d by a variety of clients to manage critic at events such as product launches, policy platform developments, reputation crisis management, and disinformation campaigns.

We are providing time series social media data relevant to selected assets. The data is extracted from Twitter, Reddit, Seeking Alpha, and Telegram.

The data provided can be split into 4 categories:

1. Data describing sentiment of social media posts
   a. Number of social media posts with bullish/bearish sentiment towards a target asset per period
   b. Number of upvotes/downvotes, likes, replies, comments, cross-posts of the posts with bullish/bearish sentiment towards target asset per period

2. Data describing activity of social media accounts
   a. Number of social media posts per period

3. Data describing engagement of social media accounts
   a. Number of likes and upvotes/downvotes per period
   b. Number of replies and comments to the posts per period
   c. Number of retweets and cross-posts per period

4. Data describing credibility of social media accounts
   a. Number of Social media posts done by accounts identified as bots/not bots per period
   b. Number of Upvotes/downvotes, likes, replies, comments, cross-posts of the posts done by accounts identified as bots/non-bots per period
   c. Number of social media posts done by accounts identified as influencers/market analysts per period
   d. Number of upvotes/downvotes, likes, replies, comments, cross-posts of the posts done by accounts influencers/market analysts per period
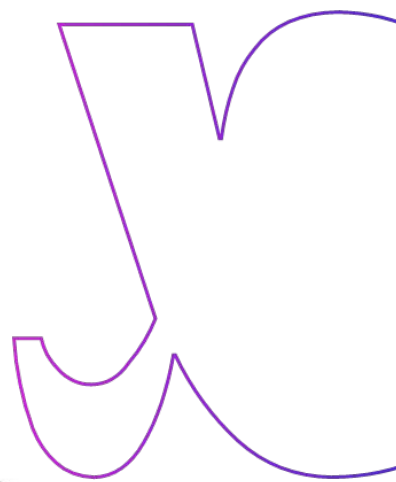
Data analytics methodology

Selection of asset-relevant social media posts:

This task is done via iterative usage of information retrieval methods such as keyword extraction and topic modelling (LDA, BERTopic, etc.). We extract the keywords for each asset that are commonly used by people. Because a person who wants to influence public opinion on an asset must provide a specific name for the target asset, such as relevant codes or common names, the keywords they choose will help us to identify them. Also, there are fine-tuned models to help us to determine the truth about the financial topics. By combining these methods and models, we can focus on the data to seek the alpha or identify critical events from different influencers.

Financial-related classification:

To filter the key samples from large amounts of posts and news, we employ one of the state-of-art NLP models (Roberta-XLM) to achieve the best performance. There were already some pre-trained models focused on the news containing traditional assets such as bonds, FX, and stocks. By using weak-supervision learning and the additional internal data related to less traditional assets like crypto (added via such techniques as pseudo-labelling), our fine-tuned classifier can achieve great accuracy and precision. This is a binary classification to predict whether the post is related to finance or not.

ZENPULSAR

Account classification

To classify an account as a bot or as an authentic user, we apply a combination of the following techniques:

- NLP-based content analysis - we employ transformer models like google MT5 or XLM-Roberta trained on bot post datasets.

- Heuristics-based features (speed of posting, statistical characteristics based on NER analysis results, etc). Those features are fed to the Support Vector machine classifier.

- The format of recent posts from the same user. Many bots have templates for different posts by putting the text together and transforming it. The model can extract features on it to improve the model.

- Analysis of network topology (bots have a different one from human accounts), specifically betweenness centrality characteristics of an account within an account network (Katz centrality, Pagerank).

To classify an account as an influencer, a market analyst, or an abnormal user, we apply a combination of the following techniques:

- NLP-based content analysis - transformer models like google MT5 or XLM-Roberta trained on influencer post datasets.

- Analysis of the account following network characteristics of an account, specifically betweenness centrality, within the account network (Katz centrality, Pagerank, Eigenvector centrality).

- Number of followers/reddit karma thresholds.

Sentiment detection:

We utilise transformer-based models (FinBert, CryptoBert and CryptoRoberta) fine tuned on our internal datasets. The model was trained on cryptocurrency and stock data collected from social media, and three classes will be output by the classifier, bearish, neutral, and bullish.

ZENPULSOR

Asset Coverage

Major commodities assets covered with new commodities added regularly.

Dataset attributes

Attribute	Type	Description	Example
time stamp	time stamp	Data and time of a datapoint	2023-02-07T00:00:00+00:00
comments	int	Number of comments	1675
likes	int	Number of likes	8598
posts	int	Number of posts	10108
reposts	int	Number of reposts	2300
code	list	List of tracked assets	Brent
account_types	list	Types of accounts posted, reposted or comments a post (bot, influencer)	Is_influencer
sentiments	list	Sentiments of a post r a comment (bullish, bearish)	Is_bullish
sources	list	Social media network (Twitter, Reddit, Telegram, Seeking Alpha)	twitter

Example of output

timestamp	comments	likes	posts	repost	code	Accunt_types	sentiments	sources
2023-02-07T00:00+00:00	1675	8598	10108	2300	Brent	Is_influencer	Is_bullish	twitter
2023-02-08T00:00+00:00	2074	10693	13207	3161	Brent	s_influencer	Is_bullish	twitter
2023-02-09T00:00+00:00	3057	21504	13357	5244	Brent	s_influencer	Is_bullish	twitter
2023-02-10T00:00+00:00	1838	8918	11186	2488	Brent	s_influencer	Is_bullish	twitter
2023-02-11T00:00+00:00	1839	8725	11617	2661	Brent	s_influencer	Is_bullish	twitter
2023-02-12T00:00+00:00	1 065	4602	7896	1169	Brent	s_influencer	Is_bullish	twitter
2023-02-13T00:00+00:00	1 085	5609	6674	1251	Brent	s_influencer	Is_bullish	twitter
2023-02-14T00:00+00:00	1 512	7150	9028	1982	Brent	s_influencer	Is_bullish	twitter

ZENPULSOR

Data quality

99 % of data consistency

Data volume

over than 0,5B data points

Country coverage

Worldwide

Delivery Format
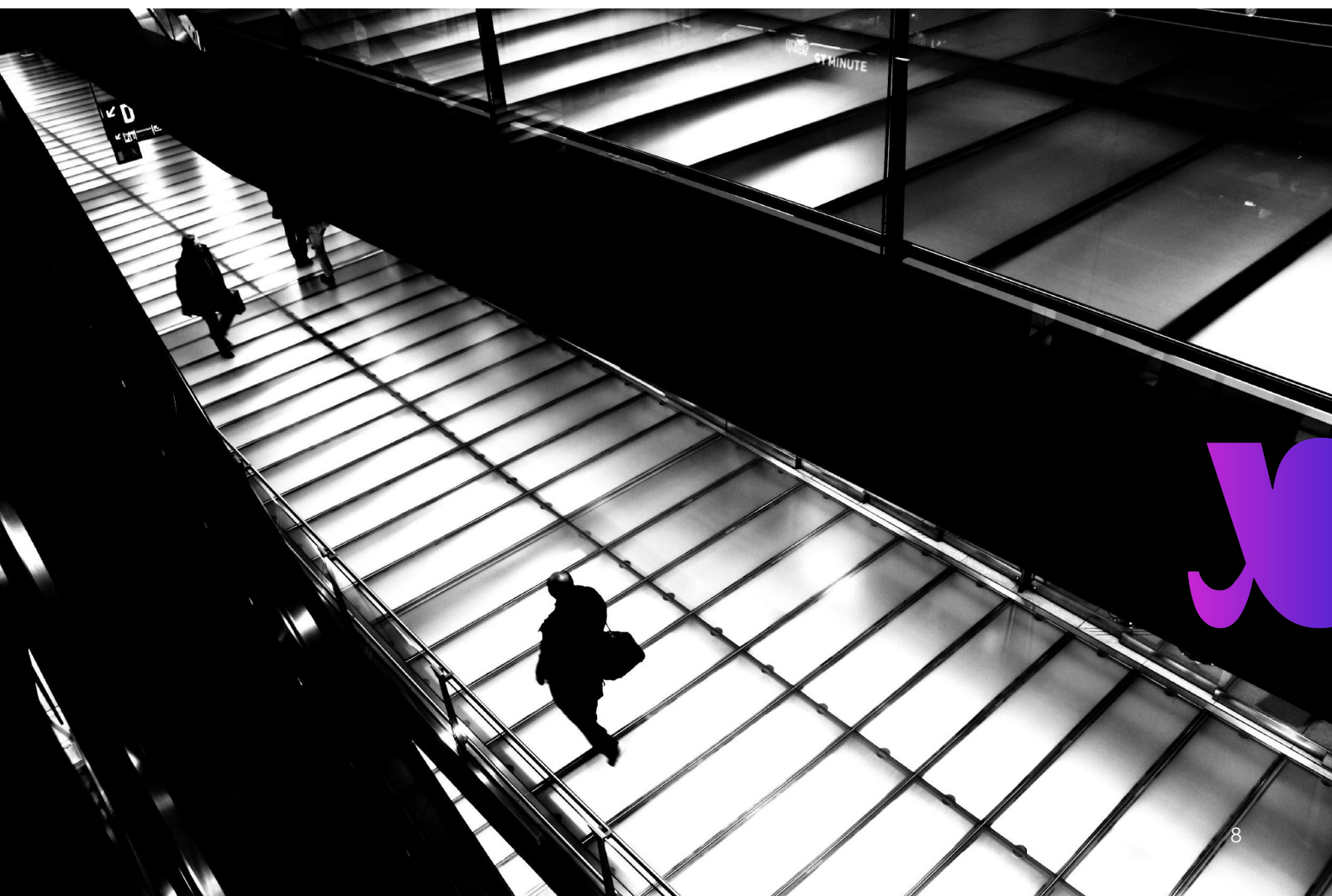
JSON, CSV

Delivery Method

REST API, Swagger available

Delivery Frequency

Hourly, Daily, Weekly, Monthly

Use cases

- Sentiment Analysis
- Hedge Funds
- Asset Management
- Quantitative Investing
- Alpha Generation

CONTACTS

ALEXANDER PISEMSKIY
Chief Executive Officer
e: ap@zenpulsar.com
m: +44 (0)7388802237

JULIEN ARTERO
Market & Sales Strategy
e: ja@zenpulsar.com
m: +44 (0)7920038238

Zenpulsar Ltd
20-22 Wenclock Road
N1 7GU London,
UK

zenpulsar.com

ZENPULSAR