ZENPULSAR

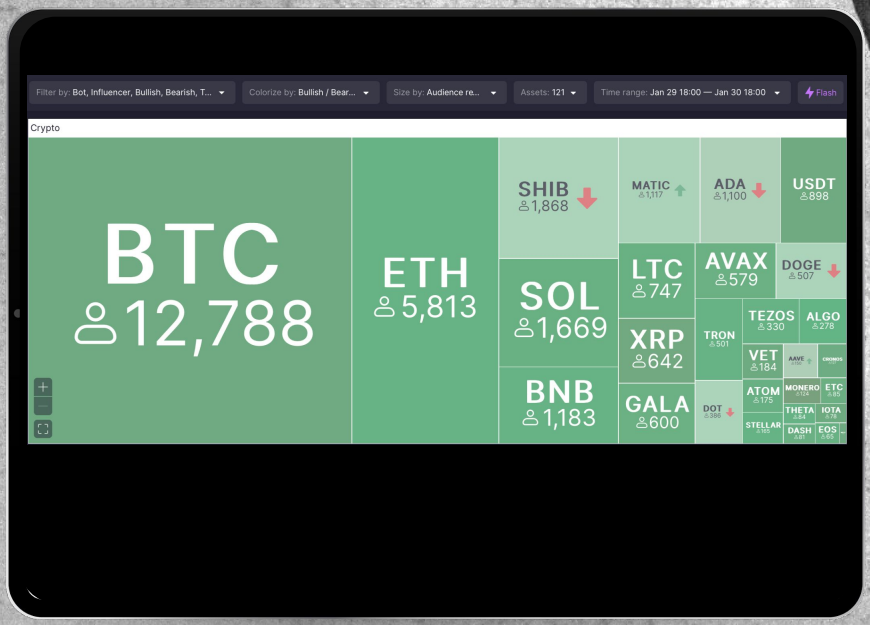
Transforming Social Media Data Into Value

Social Media Pulse Data Set

Momentum

April 2023

ZENPULSAR



Social Media Momentum Data Set

ZENPULSAR's PUMP Social Media Momentum tracks mentions of assets in social media and evaluates their popularity. This data set provides information about trending assets in Twitter, Reddit, Telegram, and Seeking Alpha, as well as how the dynamics of popularity changes among different groups of users: influencers, bots, and retail investors.

Social Media Momentum

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 analyses 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 you can spot and act on before the crowds and everyone else.

Beyond financial services, ZENPULSAR's technology is leveraged by a variety of clients to manage critical events such as product launches, policy platform developments, reputation crisis management, and disinformation campaigns.

ZENPULSAR's PUMP Social Media Momentum provides data about the popularity of monitored assets from different classes like Equities, Crypto, Commodities, FX and Fixed Income in social media platforms: Twitter, Reddit, Seeking Alpha, and Telegram.

Popularity is measured based on various metrics: audience reach, number of posts, comments, likes, and reposts. Our dataset gives the ability to rank assets not only by number of mentions, but also by the level of engagement of the audience to narratives related to the asset and sentiment.

Data set can be filtered according to the following parameters:

- Asset name/Ticker;
- Social media networks (Twitter, Reddit, Seeking Alpha, and Telegram);
- Type of accounts (bots or influencers);
- Sentiment (bullish or bearish);
- Time frame.

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 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 state-of-the-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.

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 or 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.

Asset Coverage

- Equities listed on major stock exchanges
- Major commodities
- Top capitalization cryptocurrencies
- Currency pairs
- US Treasuries
- New assets added regularly

Dataset attributes

Attribute	Type	Description	Example
asset_codes	list	list of tracked assets	AAVE
account_types	list	types of accounts posted, reposted or commented a post (bot, influencer)	is_influencer
sentiments	list	sentiment of the post or comment (bullish, bearish)	is_bullish
groups	list	Group of assets (crypto, commodities, equity, fx, fixed income)	Crypto
measurements	list	Popularity measurement (audience reach, comments, likes, post, reposts)	comments
sources	list	social media network (Twitter, Reddit, Telegram, Seeking Alpha)	twitter
start_datetime	str	Start date of the timeframe	2022-09-16T21:00:00+00:00
end_datetime	str	End date of the timeframe	2023-11-23T12:30:00+00:00
group_interval	str	Interval of data aggregation	1dœ
ticker	str	Ticker of the asset	AAVEUSDT
name	str	Asset name	AAVE
key	str		faa900b42941966a980aaa2a8a6ae560
audience_reach	float	Proprietary measurement which represents number of accounts which saw, reacted to the post or narrative	27
comments	int	Number of comments	9
likes	int	Number likes	23
posts	int	Number of posts	14
reposts	int	Number of reposts	0
asset_codes	list	list of tracked assets	AAVE
account_types	list	types of accounts posted, reposted or commented a post (bot, influencer)	is_influencer
sentiments	list	sentiment of the post or comment (bullish, bearish)	is_bullish
groups	list	Group of assets (crypto, commodities, equity, fx, fixed income)	Crypto
measurements	list	Popularity measurement (audience reach, comments, likes, post, reposts)	comments
sources	list	social media network (Twitter, Reddit, Telegram, Seeking Alpha)	twitter
start_datetime	str	Start date of the timeframe	2022-09-16T21:00:00+00:00
end_datetime	str	End date of the timeframe	2023-11-23T12:30:00+00:00
group_interval	str	Interval of data aggregation	1dœ
ticker	str	Ticker of the asset	AAVEUSDT
name	str	Asset name	AAVE
audience_reach	float	Proprietary measurement which represents number of accounts which saw, reacted to the post or narrative	27
comments	int	Number of comments	9
likes	int	Number likes	23
posts	int	Number of posts	14
reposts	int	Number of reposts	0

Example of output

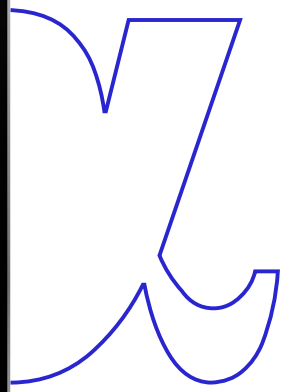
Request: query: {"asset_codes": [ ],
"account_types": ["is_influencer"],
"fields": ["is_bullish"
"is_bearish",
"is_bot",
"is_influencer",
"source"],

"measurements": ["audience_reach",
"comments",
"likes",
"post",
"reposts"],

"sentiments": ["is_bullish",
"is_bearish"],

"sources": ["reddit",],

"start_datetime": "2022-09-16T21:00:00+00:00",
"end_datetime": "2023-11-23T12:30:00+00:00",
"group_interval": null,
"meta": true }



assetcode	ticker	name	group	key	audience_reach	comments	likes	posts	reposts	is_bullish	is_bullish	is_bearish	Type
AAVE	AAVEUSDT	AAVE	Crypto	faa900b42941966a980aaa2a8a6ae560	27	9	23	14	0	true	false	true	true
ADA	ADAUSDT	Cardano	Crypto	faa900b42941966a980aaa2a8a6ae560	214	73	180	110	0	true	false	true	true
ALGO	ALGOUSDT	Algorand	Crypto	faa900b42941966a980aaa2a8a6ae560	1081	316	828	663	8	true	false	true	true
ATOM	TRXUSDT	Cosmos	Crypto	faa900b42941966a980aaa2a8a6ae560	325	111	231	182	3	true	false	true	true
AVAX	AVAXUSDT	Avalanche	Crypto	faa900b42941966a980aaa2a8a6ae560	58	18	72	31	0	true	false	true	true

Data quality

99 % of data consistency

Data volume

over than 0,5B data points

Country coverage

Worldwide

Delivery Format

JSON, CSV

Delivery Method

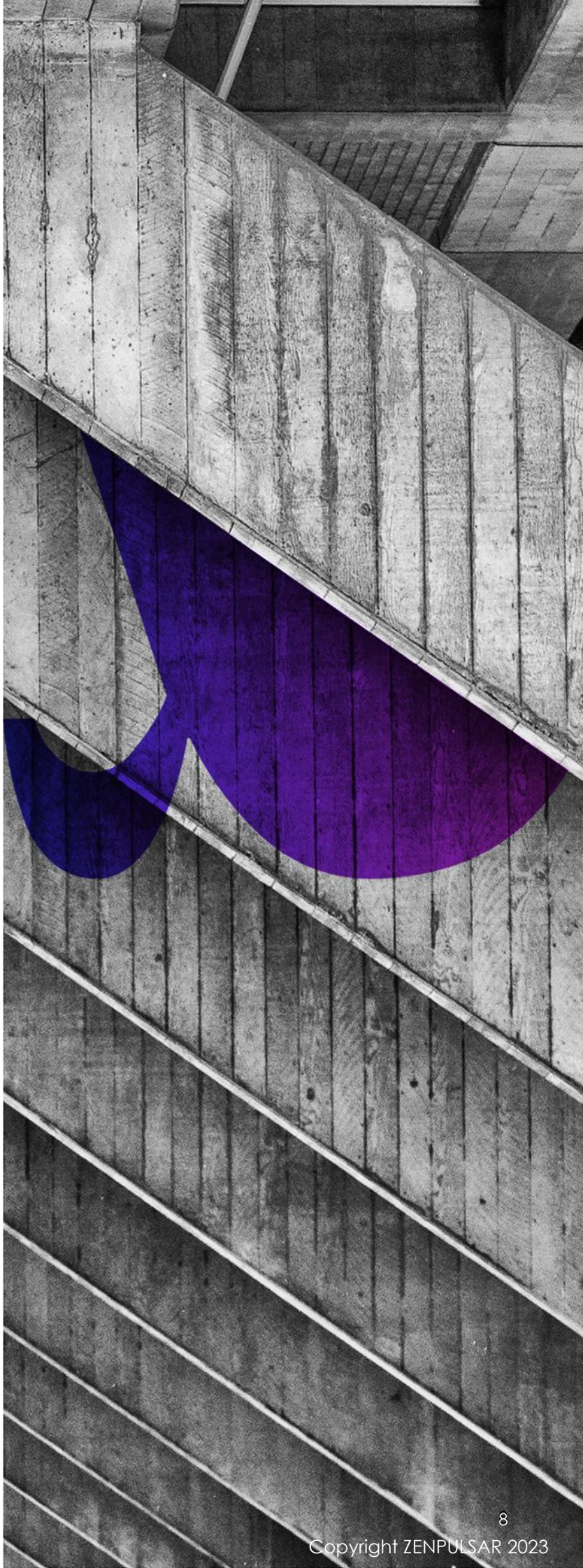
REST API, Swagger available

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