



Augmento sentiment data

Technical data description

1. General description

The Augmento data pipeline ingests posts from Twitter, Reddit, and Bitcointalk, utilizing machine learning and natural language processing to quantify sentiment for 77 crypto assets at hourly and daily intervals. The sentiment data can be retrieved via a REST API.

2. Data structure, calculation and validity

- Historic data for many assets starts around November 2016
- Timestamps are UTC
- The timestamp "2022-01-14 12:00:00" indicates that sentiment data was captured from "2022-01-14 12:00:00" to "2022-01-14 13:00:00." This data row becomes available around "2022-01-14 13:10:00" during live operation. Accessing live data requires an API key.
- The provided data reflects the count of individual social media posts. For example, there were "x" posts about crypto asset "x" labeled with the category "fear." You can aggregate this data using a sum function. The total amount of processed posts per time period is not available. If a baseline is required, it is recommended to use a moving average.
- Each social media post is analyzed by a machine learning classifier, which may assign values to multiple sentiment topics for a given post.

3. API documentation

You can find our recent REST API documentation here:
<https://api.augmento.ai/v0.1/documentation#rest-api-endpoints>

4. Sentiment topic description

A comprehensive document outlining the sentiment topics is available to subscribers (Augmento – topics vX.xlsx).

5. API key restrictions

All historical data can be accessed without an API key, except for the most recent 30 days. To explore options for obtaining an API key and accessing live data, please contact us at augmento@postera.io.

6. Best practices for implementing automated Augmento data ingestion

To ingest all data into a local database efficiently, the following process is recommended to ensure optimal processing times:

```
GET https://api.augmento.ai/v0.1/events/aggregated?source={source}&coin={coin}&bin_size=1H&count_ptr=1000&start_ptr=0&start_datetime={start_datetime}&end_datetime={end_datetime}
```

Iterate through all available sources (source=twitter, reddit, bitcointalk). Iterate through all available coins (coin=bitcoin, ethereum ...) for each source.

The start_ptr should be set to a fixed value of 0, and the count_ptr should be set to a fixed value of 1000 for every request. The only parameters that will be modified with each request are start_datetime and end_datetime, which must be formatted in ISO 8601 (e.g., 2019-04-06T00:00:00Z). The data should be queried by incrementing both the start_datetime and end_datetime by 40 days for each subsequent request, ensuring that the interval between them remains 40 days. The expected response time for these requests is typically between 1 to 3 seconds.

This API call returns hourly sentiment counts in the following format:

```
{
  "counts": [0, 2, 1 ... 0, 2, 39],
  "datetime": "2019-04-01T00:00:00Z",
  "t_epoch": 1554076800
},
```

Each position in the counts list corresponds to a specific sentiment topic in a defined order. You can find the complete list of ordered sentiment topics here: <https://api.augmento.ai/v0.1/documentation#topics>

It is advisable to perform the initial full history data ingestion using serial (non-parallel) API requests to avoid degrading API performance. To append live data to your local sentiment database, request data from the API using start_datetime=now-48h in serial requests, executed 5 minutes after each full hour. Then, add any new hourly data rows that are not already present in your database.

IMPORTANT: The most recent data row (e.g., 1h) is provisional and may change. For stable and finalized data, please refer to the 1h row preceding the most recent one. This information is applicable only to users with an API key.

Feel free to contact us if you have any further questions at augmento@postera.io.