

BÖRSE STUTTGART

March 27, 2025

1 X Solana tweets Analysis

We are going to use official twitter API to pull out Solana related posts, and then analyse them using roBERTa model and OpenAI gpt4-o model. Let's start with basic imports:

```
[60]: import tweepy
import pandas as pd
from langchain_community.callbacks import get_openai_callback
from langchain_core.messages import HumanMessage, SystemMessage
from langchain_openai import ChatOpenAI
```

Warning: Free tier of twitter API only allows pulling **100 tweets** a month, hence after the initial run we save out pulled tweets to a pickle file, for reruns and analysis.

So, to run it freshly, remove the commenter API call and load data from API rather than a dataframe. Assumption is for actual project we won't have this bottleneck, worst case we can make use of available open source library for pulling tweets, however these are unreliable and risk breakage.

```
[93]: # # Set up Twitter API credentials
# BEARER_TOKEN =
↳ "AAAAAAAAAAAAAAAAAAAAAEcV0AEAAAAAs9DtkGAVpcg0AhI6LIrhmDE350Q%3D3eGGcicaf6hxSdjkU600ndNoMxjA

# # Authenticate and connect
# client = tweepy.Client(bearer_token=BEARER_TOKEN)

# # Fetch tweets from the past 7 days
# query = "Solana -is:retweet lang:en"
# tweets = client.search_recent_tweets(query=query, max_results=10,
↳ tweet_fields=["created_at", "text", "public_metrics"])

# # Store data in a DataFrame
# data = [{"time": t.created_at, "text": t.text, "likes": t.
↳ public_metrics["like_count"]} for t in tweets.data]
# df = pd.DataFrame(data)
# df.to_pickle('solana_tweets')

df = pd.read_pickle('solana_tweets')
df.head()
```

```
[93]:
```

	time \		text	likes
0	2025-03-27 09:55:17+00:00		PLUME Coin Price Prediction, What's Next for P...	0
1	2025-03-27 09:55:12+00:00		@GamerBoom_ Airdrop-\$11M Backed by Solana &am...	0
2	2025-03-27 09:55:09+00:00		@ZygoFrog Good Job Team Zygo , base and Solana...	0
3	2025-03-27 09:55:08+00:00		Buy the dip? Bro, I've been buying dips since ...	0
4	2025-03-27 09:55:07+00:00		@GamerBoom_ Airdrop-\$11M Backed by Solana &am...	0

```
[95]: # Print number of rows and also print the tweets in full length for
↳visuaalization
print("No. of rows: ", df.shape[0])
with pd.option_context('display.max_rows', None, 'display.max_columns', None,
↳'display.max_colwidth', None): # more options can be specified also
display(df)
```

No. of rows: 10

```
time \
0 2025-03-27 09:55:17+00:00
1 2025-03-27 09:55:12+00:00
2 2025-03-27 09:55:09+00:00
3 2025-03-27 09:55:08+00:00
4 2025-03-27 09:55:07+00:00
5 2025-03-27 09:55:07+00:00
6 2025-03-27 09:55:04+00:00
7 2025-03-27 09:55:02+00:00
8 2025-03-27 09:55:00+00:00
9 2025-03-27 09:55:00+00:00

↳
↳
↳
↳
text \
0 PLUME Coin Price Prediction, What's Next for PLUME coin: Down to 15.8 Cents?
↳https://t.co/Q8xtyZ8Khr PLUME Coin Price Prediction, Analysis, News and Update
↳-27 March 2025 #cryptoanalysis #pricepredictions #solana \n #cryptocop
↳#trading #trading #plume \nGet the latest Plume c...
1 @GamerBoom_ Airdrop-$11M Backed by Solana &am
↳NVIDIA\n\n1 Sign up: https://t.co/o5f9qayWfS \n2 Do social tasks + daily
↳check-in \n3 Invite friends or download the app to get more
↳pts\n@CryptoSaiV2 @CryptoGuuz @adilk0870 @benson_btc @mominsaqib
```

2
↳
↳
@ZygoFrog
↳ Good Job Team Zygo , base and Solana to the
3 Buy the dip?
↳ Bro, I've been buying dips since BTC was \$40K. Just send it already. \n \n
↳ https://t.co/OCqmHMK0gC solana,memecoin,BITCOIN,Crypto \$QQQ \$GME \$VKTX \$FFIE
↳ \$AMC \$GWAV \$CRKN \$SPY \$GNS \$HOLO \$NVDA \$CYN \$BSFC
4 @GamerBoom_ Airdrop-\$11M Backed by Solana & NVIDIA\n\n1
↳ Sign up: https://t.co/MXIrWbifgZ \n2 Do social tasks + daily check-in \n3
↳ Invite friends or download the app to get more pts\n@beautyofhelin
↳ @isadora288881 @MartiniGuyYT @noriochan_game @TheDustyBC
5 The market is always searching for the best price. Find the
↳ hidden gems before the herd does. @JupiterExchange is building something
↳ special. #Solana #JupiterExchange #JUPtoken #Jupnet
6
↳
↳
↳
@Solana_Slot Gm boss
7 @GamerBoom_ Airdrop-\$11M Backed by Solana &
↳ NVIDIA\n\n1 Sign up: https://t.co/eNejz56srf \n2 Do social tasks + daily
↳ check-in \n3 Invite friends or download the app to get more pts\n@dyor_100x
↳ @zenryoku_eth @gg776650 @golden_kripto @blokinfluencer
8 Every
↳ wallet gets 250k \$KING \n\nDrop your Solana \$Sol Address & RT \n\nLike
↳ & Rt pinned post https://t.co/hhyzCo9iE3
9
↳ @solana_zzz Whoa, that's some wild code! If you're into unique vibes, check
↳ out Catapult Corgi-we just dropped on Solana! Transparent community, no
↳ gimmicks. Join us\n@catapultcorgi\nwith \$CCC #memecoin!

likes

0 0
1 0
2 0
3 0
4 0
5 0
6 0
7 0
8 0
9 0

1.1 Analysis

Now that we have our tweet, we now analyze our tweets for sentiment, credibility, relevance, reliability and category. We plan to use:

- roBERTa library for sentiment analysis(state of the art for sentiment analysis, and individually suited for the task rather than general llm)
- openAI gpt 4-0 model for rest of analysis. This assigns a score between 1-5 and also reasoning for the score.

1.1.1 Warning: I have used my own personal GPT which is costly, hence the access key has been commented out to avoid unfair usage.

```
[74]: import torch
import json
from ollama import Client
from transformers import pipeline

# Initialize sentiment analyzer with roBERTa
sentiment_analyzer = pipeline(
    "text-classification",
    model="cardiffnlp/twitter-roberta-base-sentiment",
    device=0 if torch.cuda.is_available() else -1
)

# Define label mapping for the sentiment model
label_mapping = {
    "LABEL_0": "NEGATIVE",
    "LABEL_1": "NEUTRAL",
    "LABEL_2": "POSITIVE"
}

# Apply sentiment analysis
def analyze_sentiment(text):
    result = sentiment_analyzer(text, truncation=True)[0] # Single call
    return label_mapping[result["label"]], result["score"]

# Apply function and unpack into two columns
df[["sentiment", "sentiment_score"]] = df["text"].apply(lambda x: pd.
    ↳Series(analyze_sentiment(x)))
# Convert timestamp column to string format (ISO 8601)
df["time"] = df["time"].astype(str)

# Prompt for the LLM for analysis, it can be easily modified to change the
↳queries on a as-needed basis.
messages = [
    HumanMessage(
```

```

        content=[
            {
                "type": "text",
                "text": ""
                You MUST return a valid JSON object following this exact
↪schema:
                {
                    "credibility": "int between 1-5",
                    "credibility_reasoning": "string",
                    "relevance": "int between 1-5",
                    "relevance_reasoning": "string",
                    "reliability": "int between 1-5",
                    "reliability_reasoning": "string",
                    "category": "string",
                    "category_reasoning": "string"
                }
                Do NOT include extra text. Output only JSON string no
↪indentation nothing.
                ""
            },
            {
                "type": "text",
                "text": f"Text to analyze: {twc}"
            }
        ]
    )
]

chat = ChatOpenAI(model="gpt-4o", max_tokens=256, api_key = "YOUR_API_KEY")

json_array = []
for twc in df["text"]:
    with get_openai_callback() as cb:
        result = chat.invoke(messages)
        print(cb)
        print("response: ", result.content)

    # Convert JSON string to Python list
    json_obj = json.loads(result.content)
    json_array.append(json_obj)

# Print the json and convert into new df
print(json.dumps(json_array, indent=4))

new_df = pd.DataFrame(json_array)

```

Device set to use cpu

Asking to truncate to max_length but no maximum length is provided and the model has no predefined maximum length. Default to no truncation.

analyse:

Tokens Used: 342

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 149

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0019725

response: {"credibility": "2", "credibility_reasoning": "The tweet is a mention of multiple Twitter handles followed by an informal expression 'ily sis' and a link. There is no context or information to assess credibility.", "relevance": "1", "relevance_reasoning": "The tweet does not provide any substantial information or context related to specific topics or events.", "reliability": "2", "reliability_reasoning": "The tweet does not contain any verifiable information or claims, making it not particularly reliable.", "category": "Social Media Interaction", "category_reasoning": "The tweet involves a common social media interaction with mentions and an informal message, typical of personal communication or networking."}

analyse:

Tokens Used: 319

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 126

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0017425000000000001

response: {"credibility": "2", "credibility_reasoning": "The content is a social media post with multiple user tags and a URL, lacking context or verifiable information.", "relevance": "1", "relevance_reasoning": "The post lacks substantive information or relevance to broader topics as it primarily focuses on tagging users.", "reliability": "2", "reliability_reasoning": "Given the informal nature and lack of content, the reliability of this post can not be determined and is therefore low.", "category": "Social Media Post", "category_reasoning": "This is a typical social media interaction where users are tagged without further insightful content or context."}

analyse:

Tokens Used: 306

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 113

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0016125000000000002

response: {"credibility": "2", "credibility_reasoning": "The post is a personal message with no factual information to verify.", "relevance": "1",

"relevance_reasoning": "The content is personal and lacks relevance to a broader audience or significant topic.", "reliability": "2", "reliability_reasoning": "The message includes a social media link, but the information cannot be evaluated for reliability.", "category": "personal communication", "category_reasoning": "The message is a personal greeting between individuals."} analyse:

Tokens Used: 306

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 113

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0016125000000000002

response: {"credibility":2,"credibility_reasoning":"The tweet consists of tagged usernames and a shortened URL, providing no substantial information to assess credibility.", "relevance":1,"relevance_reasoning":"The tweet lacks context or content that indicates its relevance to any particular topic or event.", "reliability":2,"reliability_reasoning":"The reliability is uncertain as the tweet provides no verifiable information or sources.", "category":"social media", "category_reasoning":"The format of the tweet is typical for social media interactions, involving user mentions and links."} analyse:

Tokens Used: 337

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 144

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0019225

response: {"credibility":2,"credibility_reasoning":"The text appears to be a personal message or casual interaction rather than factual information, and includes usernames and a link which does not provide verifiable information.", "relevance":1,"relevance_reasoning":"The text lacks specific content or context related to a topical subject or issue, making it not relevant for informational purposes.", "reliability":2,"reliability_reasoning":"With no factual content to verify and its nature as a casual message, it does not stand as a reliable source of information.", "category":"Social Media Interaction", "category_reasoning":"The format and content suggest it is a social media post or tweet aimed at users tagged in the message."} analyse:

Tokens Used: 328

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 135

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0018325000000000001

response: {"credibility":"1","credibility_reasoning":"The text is a tweet

mentioning several Twitter handles and a URL, with no informative content to assess credibility.", "relevance": "1", "relevance_reasoning": "The tweet provides no useful information or context related to any specific topic, making it irrelevant to most informational needs.", "reliability": "1", "reliability_reasoning": "The tweet contains no verifiable information or sources, just social media handles and a generic message.", "category": "social", "category_reasoning": "The structure of the text and use of '@' handles indicate that it is social media content, specifically a tweet, which aligns with the social category."}

analyse:

Tokens Used: 352

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 159

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0020725

response: {"credibility": 3, "credibility_reasoning": "The text contains a list of Twitter handles and a generic message, without providing any specific information or claims that can be verified for credibility.", "relevance": 2, "relevance_reasoning": "The text appears to be a personal message to specific users on Twitter, making it only relevant to those individuals, and possibly to their followers, within the Web3 and cryptocurrency space.", "reliability": 3, "reliability_reasoning": "The tweet is a personal message and does not attempt to provide factual information. Its reliability depends on the context and relationships between the users mentioned.", "category": "Social Media Interaction", "category_reasoning": "The text is a tweet that mentions multiple users and includes a personal message, which is characteristic of social media interactions."}

analyse:

Tokens Used: 351

Prompt Tokens: 193

Prompt Tokens Cached: 0

Completion Tokens: 158

Reasoning Tokens: 0

Successful Requests: 1

Total Cost (USD): \$0.0020625

response: {"credibility": 2, "credibility_reasoning": "The tweet appears to be informal and does not provide any credible sources or evidence to support any claims, if made. It includes a link, but there is no context provided to assess its content.", "relevance": 1, "relevance_reasoning": "The tweet is primarily a mention of various Twitter users without any specific topic or substantial information, making it not particularly relevant to most discussions.", "reliability": 2, "reliability_reasoning": "The nature of the tweet is personal and informal, with no indication of verifiable information, reducing its reliability.", "category": "Social Media", "category_reasoning": "The content falls under Social Media due to its format and the use of Twitter handles to tag individuals, typical of social media interactions."}


```

analyse:
Tokens Used: 337
    Prompt Tokens: 193
        Prompt Tokens Cached: 0
    Completion Tokens: 144
        Reasoning Tokens: 0
Successful Requests: 1
Total Cost (USD): $0.0019225
response: {"credibility":2,"credibility_reasoning":"The text is a tweet with
mentions of various Twitter handles and a URL. It lacks substantiated
information or claims, making it low on
credibility.","relevance":2,"relevance_reasoning":"The text appears to be a
social tweet likely aimed at a specific audience interested in individuals
mentioned or linked content, having low relevance to broader
contexts.","reliability":2,"reliability_reasoning":"Since there is no context or
factual information, the tweet doesn't provide any reliable
content.","category":"social_media_post","category_reasoning":"The text is
characteristic of a social media interaction on Twitter, involving mentions and
a URL, typical of casual online communication."}
analyse:
Tokens Used: 318
    Prompt Tokens: 193
        Prompt Tokens Cached: 0
    Completion Tokens: 125
        Reasoning Tokens: 0
Successful Requests: 1
Total Cost (USD): $0.0017325
response: {"credibility":2,"credibility_reasoning":"This appears to be a tweet
with multiple mentions and a link, but lacks sufficient context to assess
credibility.","relevance":2,"relevance_reasoning":"The content of the tweet does
not provide a clear topical focus or relevance to any specific subject
area.","reliability":2,"reliability_reasoning":"The post includes several
Twitter handles and a shortened URL, but without context or description, its
reliability is difficult to determine.","category":"social
media","category_reasoning":"The text is from a Twitter post, which is a form of
social media."}
[
    {
        "credibility": "2",
        "credibility_reasoning": "The tweet is a mention of multiple Twitter
handles followed by an informal expression 'ily sis' and a link. There is no
context or information to assess credibility.",
        "relevance": "1",
        "relevance_reasoning": "The tweet does not provide any substantial
information or context related to specific topics or events.",
        "reliability": "2",
        "reliability_reasoning": "The tweet does not contain any verifiable
information or claims, making it not particularly reliable.",
    }
]

```

```

        "category": "Social Media Interaction",
        "category_reasoning": "The tweet involves a common social media
interaction with mentions and an informal message, typical of personal
communication or networking."
    },
    {
        "credibility": "2",
        "credibility_reasoning": "The content is a social media post with
multiple user tags and a URL, lacking context or verifiable information.",
        "relevance": "1",
        "relevance_reasoning": "The post lacks substantive information or
relevance to broader topics as it primarily focuses on tagging users.",
        "reliability": "2",
        "reliability_reasoning": "Given the informal nature and lack of content,
the reliability of this post can not be determined and is therefore low.",
        "category": "Social Media Post",
        "category_reasoning": "This is a typical social media interaction where
users are tagged without further insightful content or context."
    },
    {
        "credibility": "2",
        "credibility_reasoning": "The post is a personal message with no factual
information to verify.",
        "relevance": "1",
        "relevance_reasoning": "The content is personal and lacks relevance to a
broader audience or significant topic.",
        "reliability": "2",
        "reliability_reasoning": "The message includes a social media link, but
the information cannot be evaluated for reliability.",
        "category": "personal communication",
        "category_reasoning": "The message is a personal greeting between
individuals."
    },
    {
        "credibility": 2,
        "credibility_reasoning": "The tweet consists of tagged usernames and a
shortened URL, providing no substantial information to assess credibility.",
        "relevance": 1,
        "relevance_reasoning": "The tweet lacks context or content that
indicates its relevance to any particular topic or event.",
        "reliability": 2,
        "reliability_reasoning": "The reliability is uncertain as the tweet
provides no verifiable information or sources.",
        "category": "social media",
        "category_reasoning": "The format of the tweet is typical for social
media interactions, involving user mentions and links."
    },
    {

```

```

        "credibility": 2,
        "credibility_reasoning": "The text appears to be a personal message or
casual interaction rather than factual information, and includes usernames and a
link which does not provide verifiable information.",
        "relevance": 1,
        "relevance_reasoning": "The text lacks specific content or context
related to a topical subject or issue, making it not relevant for informational
purposes.",
        "reliability": 2,
        "reliability_reasoning": "With no factual content to verify and its
nature as a casual message, it does not stand as a reliable source of
information.",
        "category": "Social Media Interaction",
        "category_reasoning": "The format and content suggest it is a social
media post or tweet aimed at users tagged in the message."
    },
    {
        "credibility": "1",
        "credibility_reasoning": "The text is a tweet mentioning several Twitter
handles and a URL, with no informative content to assess credibility.",
        "relevance": "1",
        "relevance_reasoning": "The tweet provides no useful information or
context related to any specific topic, making it irrelevant to most
informational needs.",
        "reliability": "1",
        "reliability_reasoning": "The tweet contains no verifiable information
or sources, just social media handles and a generic message.",
        "category": "social",
        "category_reasoning": "The structure of the text and use of '@' handles
indicate that it is social media content, specifically a tweet, which aligns
with the social category."
    },
    {
        "credibility": 3,
        "credibility_reasoning": "The text contains a list of Twitter handles
and a generic message, without providing any specific information or claims that
can be verified for credibility.",
        "relevance": 2,
        "relevance_reasoning": "The text appears to be a personal message to
specific users on Twitter, making it only relevant to those individuals, and
possibly to their followers, within the Web3 and cryptocurrency space.",
        "reliability": 3,
        "reliability_reasoning": "The tweet is a personal message and does not
attempt to provide factual information. Its reliability depends on the context
and relationships between the users mentioned.",
        "category": "Social Media Interaction",
        "category_reasoning": "The text is a tweet that mentions multiple users
and includes a personal message, which is characteristic of social media

```

```

interactions."
  },
  {
    "credibility": 2,
    "credibility_reasoning": "The tweet appears to be informal and does not
provide any credible sources or evidence to support any claims, if made. It
includes a link, but there is no context provided to assess its content.",
    "relevance": 1,
    "relevance_reasoning": "The tweet is primarily a mention of various
Twitter users without any specific topic or substantial information, making it
not particularly relevant to most discussions.",
    "reliability": 2,
    "reliability_reasoning": "The nature of the tweet is personal and
informal, with no indication of verifiable information, reducing its
reliability.",
    "category": "Social Media",
    "category_reasoning": "The content falls under Social Media due to its
format and the use of Twitter handles to tag individuals, typical of social
media interactions."
  },
  {
    "credibility": 2,
    "credibility_reasoning": "The text is a tweet with mentions of various
Twitter handles and a URL. It lacks substantiated information or claims, making
it low on credibility.",
    "relevance": 2,
    "relevance_reasoning": "The text appears to be a social tweet likely
aimed at a specific audience interested in individuals mentioned or linked
content, having low relevance to broader contexts.",
    "reliability": 2,
    "reliability_reasoning": "Since there is no context or factual
information, the tweet doesn't provide any reliable content.",
    "category": "social_media_post",
    "category_reasoning": "The text is characteristic of a social media
interaction on Twitter, involving mentions and a URL, typical of casual online
communication."
  },
  {
    "credibility": 2,
    "credibility_reasoning": "This appears to be a tweet with multiple
mentions and a link, but lacks sufficient context to assess credibility.",
    "relevance": 2,
    "relevance_reasoning": "The content of the tweet does not provide a
clear topical focus or relevance to any specific subject area.",
    "reliability": 2,
    "reliability_reasoning": "The post includes several Twitter handles and
a shortened URL, but without context or description, its reliability is
difficult to determine.",
  }
}

```

```

        "category": "social media",
        "category_reasoning": "The text is from a Twitter post, which is a form
of social media."
    }
]

```

Tweets saved in 'solana_tweets.json' with correct sentiment labels & scores.

```

[97]: # visualize the new df
new_df.head()

```

```

[97]:  credibility                                credibility_reasoning relevance \
0          2  The tweet is a mention of multiple Twitter han...          1
1          2  The content is a social media post with multip...          1
2          2  The post is a personal message with no factual...          1
3          2  The tweet consists of tagged usernames and a s...          1
4          2  The text appears to be a personal message or c...          1

```

```

                                relevance_reasoning reliability \
0  The tweet does not provide any substantial inf...          2
1  The post lacks substantive information or rele...          2
2  The content is personal and lacks relevance to...          2
3  The tweet lacks context or content that indica...          2
4  The text lacks specific content or context rel...          2

```

```

                                reliability_reasoning \
0  The tweet does not contain any verifiable info...
1  Given the informal nature and lack of content,...
2  The message includes a social media link, but ...
3  The reliability is uncertain as the tweet prov...
4  With no factual content to verify and its natu...

```

```

                                category                                category_reasoning
0  Social Media Interaction  The tweet involves a common social media inter...
1      Social Media Post  This is a typical social media interaction whe...
2  personal communication  The message is a personal greeting between ind...
3      social media  The format of the tweet is typical for social ...
4  Social Media Interaction  The format and content suggest it is a social ...

```

```

[99]: # Combine sentiment analysis and llm analysis into single dataframe.
new_df = pd.DataFrame(json_array)
df = pd.concat([df, new_df], axis=1)

# Saved data to csv
df.to_csv('tweetsAnalysis.csv')

```

1.2 Conclusion and further improvements

In the above notebook we showed a POC to pull data related to Solana, and did analysis on the market sentiments through multiple angles such as sentiment, credibility, relevance, reliability. Further improvements possible:

1. Use paid account to get all tweets in the past 7 days and not be limited with free tier.
2. Custom train gpt on expected type of analysis/ output for better responses.
3. Clearly define what we mean by credibility, category etc for better tagging

1.2.1 Check the pdf version for running details of the notebook.

1.2.2 Check tweetsAnalysis.csv for results.