

# Analyzing Social Trends on Recession conditions

Kanishk Barhanpurkar  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[kbarhan1@binghamton.edu](mailto:kbarhan1@binghamton.edu)

Harshad Bhandwaladar  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[hbhandw1@binghamton.edu](mailto:hbhandw1@binghamton.edu)

Nikita Mandlik  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[nmandli1@binghamton.edu](mailto:nmandli1@binghamton.edu)

Brinda Eshwar  
Computer Science  
SUNY Binghamton University  
Binghamton, New York, USA  
[beshwar1@binghamton.edu](mailto:beshwar1@binghamton.edu)

## ABSTRACT

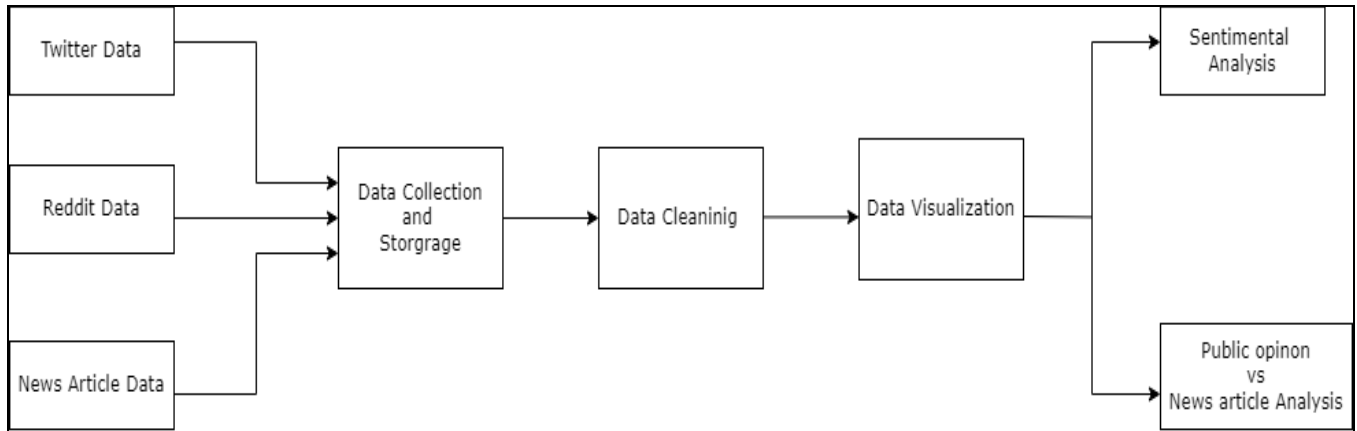
Social Media platforms are widely used for transmitting data in different formats. We can generate a lot of information on the recent trends. Additionally, public opinions can be used as a tool to predict forthcoming events. In the post-covid era, we are facing many economic challenges like the recession and other economic crises. Social media monitoring tools use to give their users insights about how the public feels in regard to their business, products, or topics of interest and can help you see how positively or negatively your brand is perceived on social media based on the tone of mentions. We are collecting data from two social networking sites, Twitter and Reddit. We are also using News Articles as another data source. Twitter is a public social networking domain and Reddit is a community-based social media platform. The primary aim of this project is to analyze the social media-associated data which predicts the conditions which lead to the Recession over time. Secondly, for the third dataset, we will evaluate the news article which contains the topic of recession and economic crisis.

## 1. INTRODUCTION

Recession is a business cycle contraction when there is a general decline in economic activity. In 2020, Covid-19 spreads exponentially in the entire world due to which many nations imposed a lockdown on the entire nation. This decision affects international trade and decreases the currency flow among various countries [1]. Several countries whose economy is dependent on the tourism sector were affected badly. Therefore, the major output of this can be seen in 2022 which results in inflation and a lack of centralized money to fulfill citizens' basic needs. Social-networking sites like Twitter and Reddit play very vital roles in understanding public opinions [2]. The number of users using social networking sites makes it mandatory to analyze the trend of current affairs happening across the globe. This project will help to compare the results obtained based on the sentimental analysis score for the public-opinion based datasets (Reddit & Twitter) with the news articles dataset (New York Times API).

## 2. DATA SOURCES

An information network made up of short messages (including photos, videos, and links) from all over the world. Twitter is a micro-blogging and social networking service where users post and interact with messages, "tweets," restricted to 280 characters. Information can be shared using photos, videos, and links. Twitter is an important social media platform because politicians, actors, and media personalities use Twitter for sharing content [3]. Reddit is a social news website and forum where content is socially curated and promoted by site members through voting. It contains a unique upvote and downvote system which helps the users to manage the public data [4]. News articles are also the major sources to gather any information regarding a particular subject across the globe. The advantage of using the news articles is proper structured language and systematic use of vocabulary which makes the analysis of text easy. Therefore, less time is required in the filtration and removing unnecessary words from the articles [5].



**Figure 1:** Data-flow diagram for social media analysis.

### 3. DATA COLLECTION (TWITTER)

In the process of data collection, we used TwitterAPI to collect Tweets from Twitter. We used a Python-based script to access the data from Twitter using Twitter API [6]. Firstly, we will import the necessary libraries for data fetching and storing operations. We have used the MySQL database for storage purposes therefore we have imported the mysql.connector library which will create a connection between MySql server and return a MySQL connection object [7]. We have provided credentials for accessing data and stored them in the database. We have created a python script such that each pull request made to Twitter API will collect a maximum of 100 tweets per hour for the keyword “Recession”. Each pull request will be created in two hours which will again fetch the tweets and store them into the database. The Twitter API is a continuously streaming API where the necessary procedures are authentication, collecting data, creating and destroying sessions. We have used Poetry, a python packaging and dependency tool which help in python libraries version control. We are collecting 16,000 tweets per week and the tweet collection will be continuous data.

### 4. DATA COLLECTION (REDDIT)

Reddit data is one of the most prominent social networking sites to gather data based on topic modeling. We have used r/Recession subreddit for the collection of data based on the user comments from Reddit API. We are collecting 300 new comments every hour and storing them into the database. We have imported requests libraries for API calls and authentication services. We have initialized a panda dataframe to store the tweets and passed the required keys from the Reddit account. We also generated

an app named app/0.0.1 which contains all the user information in the Reddit account. We have mentioned in the proposal that we are going to use RedditAPI for the data collection and we have built a Python script on the RedditAPI only. We are collecting 50,000 comments based per week based on the r/Recession data.

### 5. DATA COLLECTION (NYTIMES API)

In the proposal, we have provided that we are going to use the NewsCatcher API but we have not used the NewsCatcher API because of the pricing issues. Therefore, we are going to use the NYTimes API which is open source and provides different types of APIs to collect the data based on the user requirements. New York Times newspaper in one of the oldest and most renowned media source. We have used the Article Search API which is capable to search the news article based on the particular topic. We have used ‘Recession’ keyword to search articles for this project. In particular one API call, Python script will collect 10 articles every hour and we are storing into the database. We are collecting 1600 news articles based on economic crisis and recession conditions.

### 6. PROBLEMS FACED

We have faced a lot of difficulties in parsing the data which contains emojis and other non-alphanumeric characters. In the Twitter python script, the data collection stops as soon as it starts processing any tweet with non-alphanumeric characters. We have used type-casting and exception-handling methods provided by Python and resolved this issue. Additionally, we faced difficulty in data formatting and storing into the database. Formatting the data format such that the entire tweet must be saved in a single line so that while importing into the database the

fields will be captured properly. The reddit script would easily crash if the extracted part contains no data, therefore we built the reddit script's logic in such a way that skips the particular function.

## 7. CONCLUSION

The collected data using the Twitter API and the RedditAPI, we will be able to study and develop insights about the public opinion on the economic crisis, inflation, and the influence of upcoming-recession updates. Hence, the data will help to predict the impact of the upcoming recession on the community. Also, we will analyze the insights generated from the public opinionated datasets (Reddit and Twitter) and news articles (collected from New York Times API).

## 8. FUTURE SCOPE

In the future implementation, we are looking forward to clean and filter the data and generate some initial insights using data visualization techniques. Additionally, we will be brainstorming about the tech-stack and procedures that can be used for the data analysis.

## REFERENCES

- [1] Feyisa, H. L. (2020). The World Economy at COVID-19 quarantine: contemporary review. *International journal of economics, finance and management sciences*, 8(2), 63-74
- [2] Mariolis, T., Rodousakis, N., & Soklis, G. (2021). The COVID-19 multiplier effects of tourism on the Greek economy. *Tourism economics*, 27(8), 1848-1855.
- [3] Twitter Glossary.  
<https://help.twitter.com/en/resources/glossary>
- [4] Reddit Glossary Documentation.  
<https://www.reddit.com/r/TheoryOfReddit/wiki/glossary>
- [5] Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3), 171-188.
- [6] TwitterAPI Documentation.  
<https://developer.twitter.com/en/docs/twitter-api>
- [7]. MySQL Documentation  
<https://dev.mysql.com/doc/connector-python/en/connector-python-example-connecting.html>
- [8] RedditAPI Documentation.  
<https://www.reddit.com/dev/api/>