

README

HATE BURST

Information Retrieval Project

Parul Jain (MT21064)

Srishti Sharma (MT21098)

Lovepreet Singh (MT21044)

Syed Mohd Gulam Baquer (MT21100)

Kanak Panwar (MT21037)

About the Project

Deployed At: <http://20.239.142.79/>

HateBurst uses text classification machine learning models to identify if a given text is hateful or not. HateBurst implements several wrapper features, that use this classifier at the code for their implementation. Such as:

Twitter User Analytics: This allows the user to analyze the profile of any Twitter user by taking in just the username of the Twitter profile. The insights presented using the charts and result helps to identify, if Twitter user in general posts lots of hateful content or not, the average number of likes a hateful tweet receives vs a non-hateful one helps to identify the nature of the people who interact with that user. Other insights such as the most common words used and hashtags used in hateful tweets help identify the topic/idea that the user usually talks about, location helps to identify the locality of the user's tweets and relevance of the tweet with respect to a location.

Twitter Trends Analytics: It brings very similar results to TwitterUserAnalytics, just that instead of analyzing a single person, a trending topic could be analyzed, on the whole, if a trending topic is something for which usually hate tweets are posted, this can help stop spreading of fake news as well since if some fake news gets trending and receives lots of hateful tweets, the trend can be manually verified and be stopped from spreading further. It also provides a list of currently trending topics in India, for the user to analyze.

Hate Check: It allows the user to check whether a video, audio, or text is hateful or not, there are several features in video hate to check as well, such as checking if a video or audio file contains hateful content.

Apart from these, some features are also exposed as free-to-use REST API Services, that could be easily used by any developer or person in their projects to identify and remove hateful content from their platform.

How to Set Up the Project for Development

1. Use a Ubuntu OS or Ubuntu WSL
2. Install and configure git.
3. Install Python 3.7+
4. Install pip
5. Install virtualenv python library

```
$ sudo pip3 install virtualenv
```

6. Clone the Project into '/var/www/<here>'

```
$ sudo git clone  
https://github.com/parulj2804/Group\_20\_IR\_Project.git
```

7. Rename Group_20_IR_Project as 'hate_burst'

```
$ sudo mv Group_20_IR_Project/ hate_burst
```

8. Change mode of the project directory to 755 (access,execute by everyone + write by owner)

```
$ sudo chmod 755 hate_burst/ -R
```

9. Change Ownership to \$User

```
$ sudo chown -R $USER:$USER hate_burst
```

10. Open Project Directory i.e., 'hate_burst'

```
$ cd hate_burst
```

11. Create virtual environment

```
$ virtualenv env
```

12. Activate virtual Environment

```
$ source ./env/bin/activate
```

13. Install Dependencies

```
$ pip install youtube-dl==2021.12.17
$ pip install -r requirements.txt
```

14. Run the flask Application

```
$ python app.py
```

15. Open the website in browser localhost:5000

Dependencies:

1. absl-py==1.0.0
2. apiclient==1.0.4
3. astunparse==1.6.3
4. cachetools==5.0.0
5. certifi==2021.10.8
6. charset-normalizer==2.0.12
7. click==8.1.2
8. colorama==0.4.4
9. decorator==4.4.2
10. Flask==2.1.1
11. flatbuffers==2.0
12. gast==0.5.3
13. google-api-core==2.7.2
14. google-api-python-client==2.45.0
15. google-auth==2.6.5
16. google-auth-httplib2==0.1.0
17. google-auth-oauthlib==0.4.6
18. google-pasta==0.2.0
19. googleapis-common-protos==1.56.0
20. grpcio==1.44.0
21. gunicorn==20.1.0
22. h5py==3.6.0
23. httplib2==0.20.4
24. idna==3.3
25. imageio==2.17.0
26. imageio-ffmpeg==0.4.7
27. itsdangerous==2.1.2
28. Jinja2==3.1.1
29. joblib==1.1.0
30. keras==2.8.0
31. Keras-Preprocessing==1.1.2
32. libclang==13.0.0
33. Markdown==3.3.6
34. MarkupSafe==2.1.1
35. moviepy==1.0.3
36. nltk==3.7
37. numpy==1.22.3
38. oauthlib==3.2.0
39. opt-einsum==3.3.0

40. pafy @ git+https://github.com/Cupcakus/pafy@45f0deb067bf7c420cdf83a0529fd5274c12de18
41. pandas==1.4.2
42. Pillow==9.1.0
43. proglol==0.1.9
44. protobuf==3.20.0
45. pyasn1==0.4.8
46. pyasn1-modules==0.2.8
47. pydub==0.25.1
48. pyparsing==3.0.8
49. python-dateutil==2.8.2
50. pytube==12.0.0
51. pytz==2022.1
52. regex==2022.3.15
53. requests==2.27.1
54. requests-oauthlib==1.3.1
55. rsa==4.8
56. six==1.16.0
57. SpeechRecognition==3.8.1
58. tensorboard==2.8.0
59. tensorboard-data-server==0.6.1
60. tensorboard-plugin-wit==1.8.1
61. tensorflow==2.8.0
62. tensorflow-io-gcs-filesystem==0.24.0
63. termcolor==1.1.0
64. tf-estimator-nightly==2.8.0.dev2021122109
65. tqdm==4.64.0
66. tweepy==4.8.0
67. typing_extensions==4.2.0
68. uritemplate==4.1.1
69. urllib3==1.26.9
70. Werkzeug==2.1.1
71. wrapt==1.14.0
72. youtube-dl==2021.12.17

How to Use the WebProject

Home Page:



Click on 'Our Services' to go to the services page and use the website.

Services Page:

Hate Burst



Welcome to HateBurst
Towards a Hate-Free World

HATEBURST

```
graph TD; HB[HATEBURST] --> TUA[Twitter User Analytics]; HB --> TT[Twitter Trends Analytics]; HB --> HC[Hate Check]; HB --> APIs[APIs]; HC --> Text[Text]; HC --> Audio[Audio]; HC --> Video[Video]; Text --> THC[Text Hate Check]; Audio --> TUA[Twitter User Analytics]; Video --> TT[Twitter Trends Analytics];
```

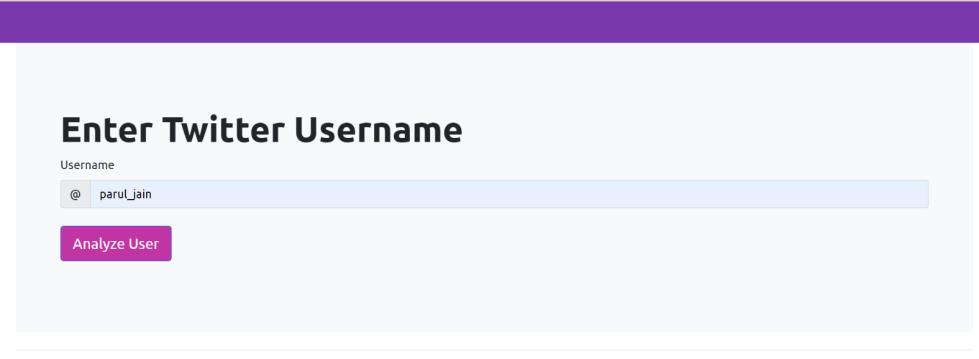
HateBurst uses text classification machine learning models to identify if a given text is hateful or not. HateBurst implements several wrapper features, that use this classifier at the code for their implementation. Such as:

- **Twitter User Analytics:** Allows the user to analyze profile of any twitter user by taking in just the username of the twitter profile. The insights presented using the charts and result helps to identify, if a twitter in general posts lots of hateful content or not, the average number of likes a hateful tweet receives vs a non-hateful one helps to identify the nature of the people who interact with that user. Other insights such as the most common words used and hashtags used in hateful tweets help identify the topic/idea that the user usually talks about, location helps to identify the locality of the user's tweets and relevance of the tweet with respect to a location.
- **Twitter Trends Analytics:** It brings very similar results to TwitterUserAnalytics, just that instead of analyzing a single person, a trending topic could be

This page displays an overview of the entire project. Click on the button on the right top corner to view the list of services offered by the web app.

Twitter User Analytics:

Hate Burst



Enter Twitter Username

Username

@ parul_jain

Analyze User

© 2022 Parul Jain

Enter a Twitter username without the '@' and click on analyze user to obtain various analytics on the same page, it may take up to a few seconds to generate the results. The results would be visible as follows:

Name: Parul Jain

Tweet	Hateful or Not
Year-End's Development check-@view https://t.co/0M3d-0nB7g	No Hate
Happiness is impossible if you are Admitted	No Hate
Count your blessings Name them one by one silently, just want more from life	No Hate
The shapes of our past define our future life@stevejobs	No Hate
After a tiring today, sleeping for hours to another tiring weekend, need to practice a little to get there ready	No Hate
Christmas again	No Hate
Everyone has one can go back yesterday and start a new beginning, but anyone can start today and make a new ending. #motivationforyou	No Hate
Everyone is not alike but unforgettable everyday, we're not perfect. All the good will is quite due to common sense	No Hate
Christmas again	No Hate
Challenge accepted!! Shared the happiness #100daysday https://t.co/0M3d-0nB7g	Hate And Abuse
Happy New Year #2014	No Hate
Not a white Christmas but still enough to feel like #ChristmasAlways#	No Hate
AlwaysChristmas everyone	No Hate
Want a white Christmas	No Hate
Cheers to the festive weekend	No Hate
Christmas is indeed magical	No Hate
The joy of a Christmas is really lovely	No Hate
Using the new year's day to bring in a new beginning everyone deserves a happy	No Hate
Want a white Christmas	No Hate
Looking forward to a beautiful day #goodmorning everyone	No Hate
#beautiful #wishes #happiness #everyday #morn https://t.co/0M3d-0nB7g	No Hate
#beautiful #sunrise #holiday #christmas #happiness	No Hate
#christmas #yayyy	No Hate
#blue #white #family	No Hate
And just like that we are in our home	No Hate
Instagram again, #christmas especially when I need all the sleep I can get	No Hate
Always Friday	No Hate
Wish the best	No Hate
Still working...	No Hate
Christmas time must be great shelter... Running out of patience now... #christmasiscoming	No Hate
Are we there yet?	No Hate
Waiting	No Hate
More's here, just 1 more...	No Hate
No one can go back yesterday and start a new beginning, but anyone can start today and make a new ending. #morn #christmas #goodmorning	No Hate
Waiting	No Hate
Always Friday, #christmas	No Hate
Always #christmas, #christmas for everyone	No Hate
will #christmas #christmas	No Hate
Beautiful #autumn #day #christmas	No Hate
Always #christmas	No Hate
And no money, no car, no house @Please #give	No Hate
Don't judge each day by the things you see, but the seeds you plant - Robert Lewis Stevenson #goodmorning everyone @. #have a beautiful day	No Hate
And it's going through my day in #christmas	No Hate
#goodmorning everyone #looking forward to an action packed day	No Hate
Always #christmas #christmas	No Hate
#christmas #christmas #christmas	No Hate
Why is the grass always greener on the other side	No Hate
Using the other items	No Hate
Good morning everyone @	No Hate
On happy day @	No Hate

Most Common Words in Hate Tweets

Challenge
accepted!!
Shared
the
happiness
#100daysday
https://t.co/0M3d-0nB7g

Most Common Locations in Hate Tweets

Sydney, Australia

HashTags in Hate Tweets

#100daysday

Legend: Hate NonHate

Sydney, Australia	1.00
-------------------	------

Twitter Trends Analytics:

Enter a Topic/Hashtag or Choose one of the Trending Hashtags from below

Enter Hashtag/Topic

Search Tweets Since: 24-03-2022

Number of Tweets to Fetch: 200

#TheRightSpiritualPath #stopHindlimposition #AjayDevgn National Language

#ACVInParallelUniverse #CTVsSRH Sant Rampal Ji Maharaj Vimal

100M MOMENTS WITH V Sudeep Telugu Hindi and English Malayalam

Cutka best wishes to eid hojayegi Varun Chakravarthy Jumla Kannada

13 YRS OF ITV KING SHAHEER Steven Y Khatri Manhandled SSR #3YrsOfSivaangi

#15YearsForSouthiHPokiri #HappyBirthdaySamantha #QOOZ6ProSG

#2DaysToDhaskadTrailer #ChristmasTree100M #VaccinationForKids #MannerlessCM

#VendhuThenindhathuKaadu #MoonKnight #BecharOutNow #Corbevax

#CabinetDecisions #SilambarasanTR #12thMan #कमीशनडॉ_कारेस

#FuelPrice #Mohanlal #HeroFan2AdvanceBooking #wednesdaythought

#Avatar2 #Xiaomi12Pro #NokiaG21 #Thalapathy67 #GotYourBack

#NationWithAmanChopra #NTR30 #UniformCivilCode #ShaheerSheikh

Analyze Trend

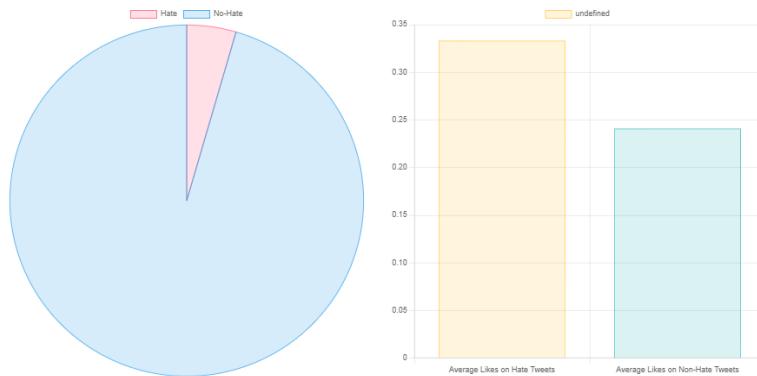
© 2022 Penul Jain

Either select one of the trending hashtags/trends on Twitter in India or enter any query hashtag/trend. Pick the date from when the tweets should be considered. Also, select the number of tweets to fetch, a maximum of up to 900 tweets can be fetched in 24 hours. The results would be displayed after waiting for a few seconds.

User	Tweet	Hateful or Not
panvith	RT @buddhikajames: hindi is not our national language, we stand with kiccha sudeep sir #stopHindilimposition https://t.co/EmNF4lVZ7M	No Hate
👉 (நாய்யு)- NAIDU 🙏	RT @urstanaysuriya: Well said sir ! Lotsa respect #KicchaBOSS 🙏 #KicchaSudeep #Kannada #AjayDevgn #VikrantRona #stopHindilimposition https://...	No Hate
guru palegara	RT @Puneethbs_Gowda: Every Kannadiga, Tamilian, Telugu, Malayali, Bengali, Assamese....should support @KicchaSudeep on this. Seems like t...	No Hate
Kishor	RT @RjSanketh: கீங்கி ராஷ்டிர சாங்கெட்டால் அதை கீங்கியினர் ஸுடிதோ கீங்கி சாங்கெட்டு வார கீங்கி இருப்பது. #stopHindilimposition #HindisNotANationalLanguage #...	No Hate
mahantesh H IN	RT @IndianM73645975: Is hindi is our national language or it's a regional language? Whom do you guys support in this matter Like - #AjayDe...	No Hate
noorullah hussain	RT @Mueen_magadi: Ohh, Bollywood actors also have the spine to speak in favour of #Hindi language which is not a National issue. Someone p...	No Hate
Raghu Gowda	RT @Puneethbs_Gowda: Every Kannadiga, Tamilian, Telugu, Malayali, Bengali, Assamese....should support @KicchaSudeep on this. Seems like t...	No Hate
S	RT @NarenDon: WhatsApp Star 😊 India is known for Diversity. No one language is National but all languages are National. To release a movie...	No Hate
👉 (நாய்யு)- NAIDU 🙏	RT @NarenDon: WhatsApp Star 😊 India is known for Diversity. No one language is National but all languages are National. To release a movie...	No Hate
ATMAN	RT @Puneethbs_Gowda: Every Kannadiga, Tamilian, Telugu, Malayali, Bengali, Assamese....should support @KicchaSudeep on this. Seems like t...	No Hate
Don Breadman 🍞	RT @ajavgal: Article 343 Official language of the union shall be Hindi in Devanagari script Why only Hindi? r v not paying tax? amend 343...	No Hate
Shambulingesh Hanchinmani	RT @urstanaysuriya: Well said sir ! Lotsa respect #KicchaBOSS 🙏 #KicchaSudeep #Kannada #AjayDevgn #VikrantRona #stopHindilimposition https://...	No Hate
Aakash	First of all, No one is imposing Hindi on Anyone. Stop Spreading your Fake Propaganda against Hindi Language Those saying	No Hate

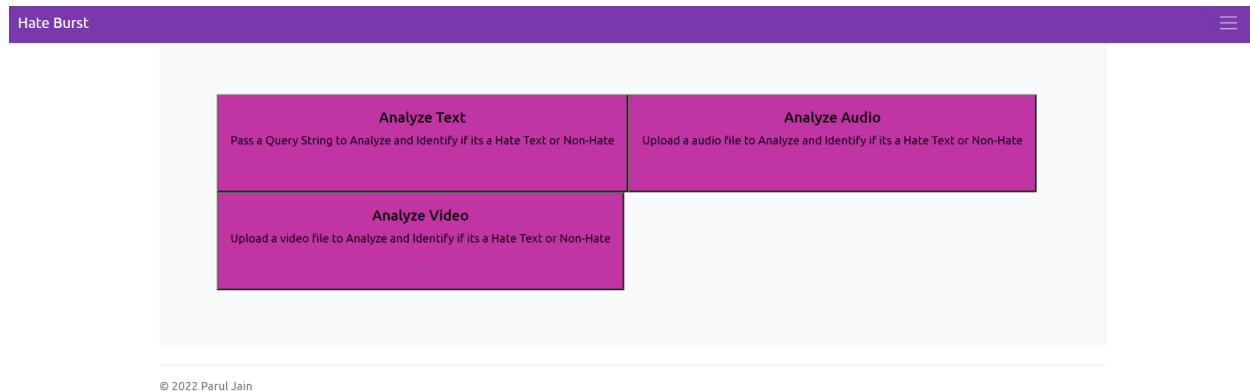
The tweets are hyperlinks to the original tweets, from where they can be reported easily.

Most Common Words in Hate Tweets	Most Common Locations in Hate Tweets	HashTags in Hate Tweets
RT #stopHindlimposition in to and their own all language +	Karnataka, India ಕರ್ನಾಟಕ	Kannada kichhasudeep stopHindlimposition AjayDevgn sudeep NationalLanguage stopHindlimposition stopHindlimposition HindiWasNeverNationalLanguage KicchaSudeep stopHindlimposition VimalPanMasala VikrantRona KicchaSudeep stopHindlimposition VimalPanMasala VikrantRona



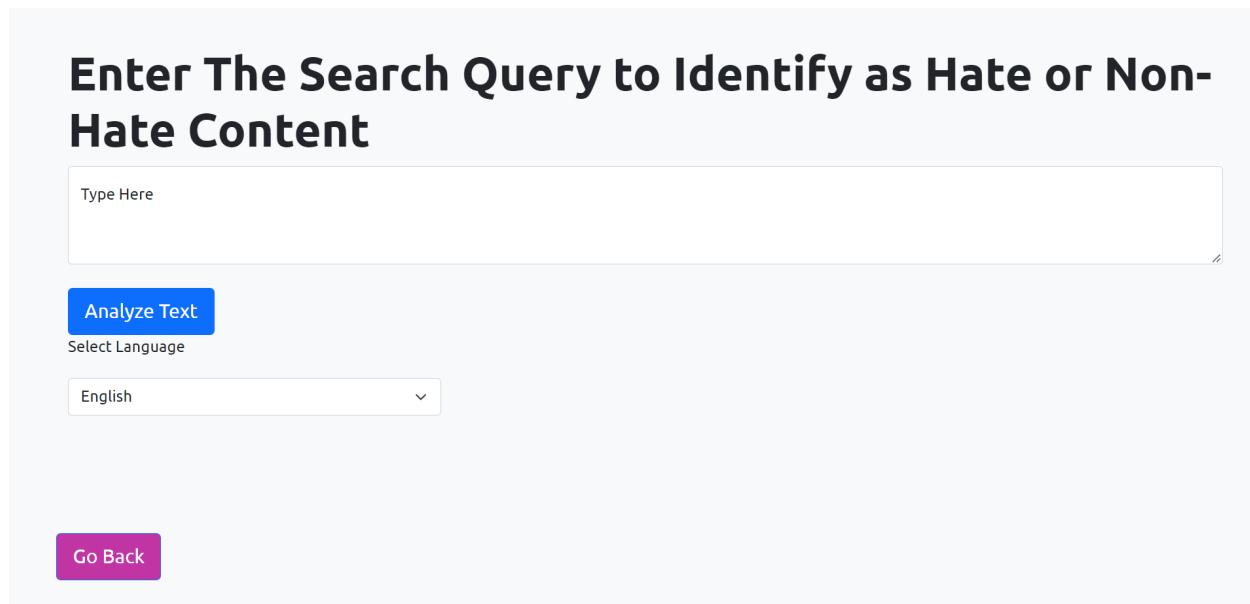
© 2022 Parul Jain

Hate Check Page:



The screenshot shows a user interface for a 'Hate Check' application. At the top, a purple header bar contains the text 'Hate Burst' on the left and a three-line menu icon on the right. Below the header are three rectangular buttons with rounded corners, each in a light gray box. The first button on the left is labeled 'Analyze Text' and 'Pass a Query String to Analyze and Identify if its a Hate Text or Non-Hate'. The second button on the right is labeled 'Analyze Audio' and 'Upload a audio file to Analyze and Identify if its a Hate Text or Non-Hate'. The third button at the bottom is labeled 'Analyze Video' and 'Upload a video file to Analyze and Identify if its a Hate Text or Non-Hate'. At the very bottom of the page, there is a small, thin horizontal line with the copyright text '© 2022 Parul Jain' centered on it.

Select any one of the three types of classifications, i.e., text, audio or video.



The screenshot shows a 'Text Analysis' page. The main title is 'Enter The Search Query to Identify as Hate or Non-Hate Content'. Below the title is a text input field with the placeholder 'Type Here'. Underneath the input field is a blue button labeled 'Analyze Text'. Below the button is a label 'Select Language' followed by a dropdown menu set to 'English'. At the bottom left of the page is a purple button labeled 'Go Back'.

For text classification, enter the text, select the language and click on Analyze Text, which would tell if the text is hateful or not.

Analyze Audio File to Identify if Hateful or Non-Hateful

Upload an Audio File (in .wav format only)

Choose File No file chosen

Analyze Audio

Go Back

To analyze the audio file, upload any audio file in .wav format, and click on Analyze Audio to identify it as hateful or non-hateful.

Analyze Video File to Identify if Hateful or Non-Hateful

Option-1: File Upload

Upload an Video File (in .mp4 format only)

Choose File No file chosen

Analyze Uploaded Video

Option-2: Youtube Video

Enter a Youtube Video URL

Analyze Youtube Video

Option-3: Youtube Search

Enter a Query to Search on Youtube

Enter Number of Results to Evaluate

1

Search Youtube Videos

Go Back

To identify the hatefulness of a video, the user can either upload an mp4 file or give a link for a youtube video or search anything on youtube and select the number of videos to analyze in the results of the search query.

How to Use the APIs

1. Install Postman from <https://www.postman.com/downloads/>
2. Use it to test the following described APIs.

API Descriptions:

GET	/api/service_classify_text
Description	Takes a query string, and identifies it as either Hateful or Non-Hateful.
Parameters	string: query
Response Body	{ prediction: Whether the Query string is Hateful or Non-Hateful, response_code: Code of Response, response_status: Message of Response, }
Errors	
900	OK
901	No argument provide
Example	/api/service_classify_text?query="you should die"

GET	/api/twitter_user_analytics/{attribute}
Description	Takes a twitter username, and provides several insights(attributes) about the posts by that user.
Parameters	
string: username	Twitter username of the Person
Attributes	
counts_pos	Number of Non-Hateful Tweets
counts_neg	Number of Hateful Tweets
tweet_count	Number of Tweets Analyzed
average_likes_hate	Average Number of Likes on the User's Hateful Tweets that are analyzed
average_likes_no_hate	Average Number of Likes on the User's Non-Hateful Tweets that are analyzed
hashtags_hate	Hashtags used in Hateful Tweets
most_common_words_hate	Most Common Words in Hateful Tweets
most_common_locations	Most Common Location in Hateful Tweets
dataframe	The python pandas dataframe of the tweets with several attributes
Response Body	{ (attribute): The value of that attribute, response_code: Code of Response, response_status: Message of Response, }
Errors	
900	OK
901	No argument provide
902	Username Doesn't Exists
903	Invalid Attribute
Example	<code>/api/twitter_user_analytics/most_common_words_hate?username="parul_jain"</code>

GET	/api/twitter_trends_analytics/{attribute}
Description	Takes a twitter trend(hashtag), and provides several insights(attributes) about the posts with that trend(hashtag).
Parameters	
string: trend	Twitter trend(hashtag)
string: date	Date since when the tweets should be analyzed, in strict format "yyyy-mm-dd"
string: count_tweets	Number of tweets that must be analyzed, should be greater than or equal to 1.
Attributes	
counts_pos	Number of Non-Hateful Tweets
counts_neg	Number of Hateful Tweets
tweet_count	Number of Tweets Analyzed
average_likes_hate	Average Number of Likes on the User's Hateful Tweets that are analyzed
average_likes_no_hate	Average Number of Likes on the User's Non-Hateful Tweets that are analyzed
hashtags_hate	Hashtags used in Hateful Tweets
most_common_words_hate	Most Common Words in Hateful Tweets
most_common_locations	Most Common Location in Hateful Tweets
dataframe	The python pandas dataframe of the tweets with several attributes
Response Body	{ (attribute): The value of that attribute, response_code: Code of Response, response_status: Message of Response, }
Errors	
900	OK
901	No argument provide
903	Invalid Attribute
904	Date Argument Missing
905	Tweet Count Argument Missing
906	Tweet Count Invalid
Example	/api/twitter_trends_analytics/most_common_words_hate?trend=ipl&date=2022-04-13&count_tweets=10

Conclusion

The project can be very useful in identifying and eliminating hateful content over the internet. Though there are certain limitations such as limited file format support, this is something that could be worked upon. Another limitation is the number of tweets that can be extracted.

