Social Media Analysis using Natural Language Processing Techniques

by

Jyotika Singh

Where there is text, there is a potential to derive meaningful insights and understanding using NLP.

Natural Language Processing on Social Media Language is Challenging!

Social Media Data

- Post/Video Statistics
- Post/Video description and other text metadata
- Audience comments

Text data, if available, opens the doors to many NLP opportunities. Some popular Natural Language Processing (NLP) tasks include the following.

- Named Entity Recognition
- Keyphrase Extraction
- Unigrams/Bigrams/Trigrams Analysis
- Tokenization
- Part-of-speech Tagging
- Lemmatization & Stemming
- Word Sense Disambiguation
- Topic Modeling
- Sentiment Analysis
- Text Summarization

Applications

- Time Window Analysis (comparing analysis between different time periods)
- Analytics and Intelligence (Trend Identification, Story Telling)

Accessing YouTube data via the YouTube API

Official YouTube Data API
https://developers.google.com/youtube/v3

Getting Started Requirements: Follow API documentation to register and enable a project and generate an API key.

Gotchas: Rate limits: API key comes with a daily rate limit that limits the number of requests that can be made to the YouTube API.
Error Handling: The API throws an error if one tries to access a video, comment or channel that was set to private by the owner. This can be an issue if one is running requests in a loop and can cause premature termination of the user's script.

pyYouTubeAnalysis library: YouTube API requests with error handling; text cleaning, keyphrase extraction; named-entity recognition; automatic report generation.

* Use the data and API in compliance with YouTube's Terms and Services

Insightful fields on YouTube

Text fields: Video title, description, tags; Channel title, description; Comment text.

Statistics: Video view, like, dislike and comment counts

Text Diversity on Social Media

Natural language on social media consists of free form text; no rules around grammar, capitalization, abbreviation, or writing style apply. Human language is ever evolving as new and popular abbreviations, topics and terms develop.

Languages
Topics
Writing styles

With the presence of content from a diverse set of topics, the terms used in text descriptions can vary vastly from something casual and informal to something technical, scientific, or formal. That is often layered in with individual writing styles and different languages from different parts of the world.

"Uh r a very gud teacher... must say uh make topics very esy... I learned constitution from ur video only... Thank uh for these videos"

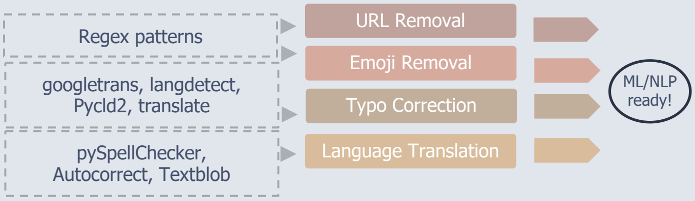
"THANK YOU FOR HELPING HIM🥺❤️"

"God bless you beautiful dog🥺🥺💗💗"

"❤️🤗"

"ਮੈਨੂੰ ਨਹੀਂ ਲਗਦਾ ਕਿ ਇਹ ਵਿਅੰਜਨ ਕੰਮ ਕਰਦਾ ਹੈ"

Text Cleaning Techniques

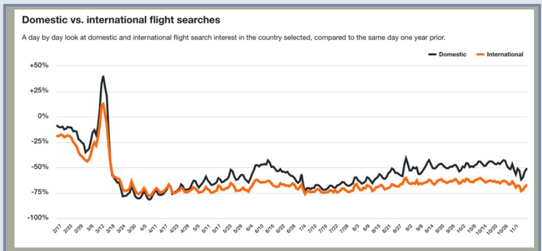


Regex patterns → URL Removal
googletrans, langdetect, Pycld2, translate → Emoji Removal
pySpellChecker, Autocorrect, Textblob → Typo Correction → Language Translation

ML/NLP ready!

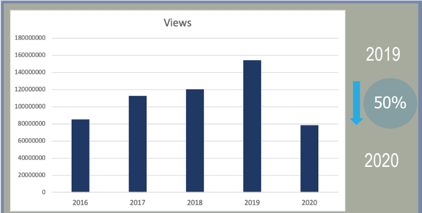
Influencing Factors and Trend Analysis

Statistical user engagement trends

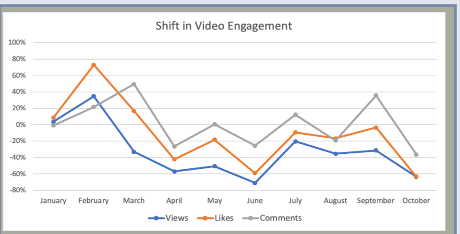


Flight searches in 2020 experienced a severe reduction after Feb-March as can be seen in this 2019 vs 2020 chart.

Search on YouTube for "travel vlogs" for 2016-2020.



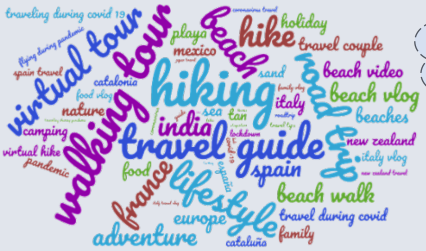
Engagement with "travel vlogs" on YouTube increased between 2016 and 2019 and then dropped off by 50% in 2020.



2019 vs 2020. A strong correlation is observed between flight search trends and engagement with travel vlogs on YouTube.

Content patterns using keyphrase extraction and named-entity recognition (NER)

The content created in 2020 was highlighting topics such as hiking and road trips where the observation of social distancing is possible, rather than crowded events.



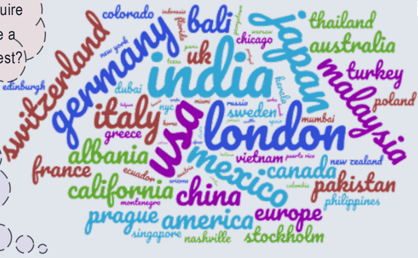
Comments held questions about COVID test requirements and travel ban lifts on several different locations across the globe.

I loved the tour looked like you guys had fun. im going there next week, how long ago were you there and were there lots of restrictions and closing due to covid

Did they require you to have a prior covid test?

Great video man, this place looks amazing. I have never been to Iceland, would love to visit some day. Honestly can't wait for the lockdown to be lifted so I can start travelling again. Thanks for sharing your experience. :)

The mention of location names found in comments hold sync with the timeline of travel ban lifts during the summer and fall of 2020.



Trending video creators in 2020

BeachTuber – Different beaches across Europe

4K Walk – Walking tours across Europe and America

Euro Trotter – Travelling all over Europe

Beach Walk – Beaches across Europe and America

DesiGirl Traveller – Travelling all over India

Tools: *pyYouTubeAnalysis for data collection, keyphrase extraction (NLTK-based) and named-entity recognition (SpaCy); wordcloud for word plots; matplotlib for statistical plots.*