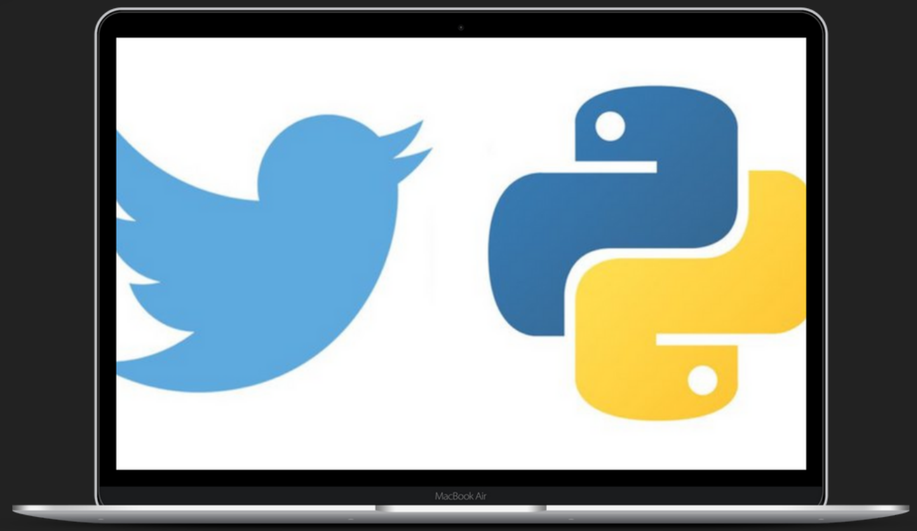
IN NAVY BLUE, WE TRUST

TWEETER INTERIM PROJECT REPORT



BRIAN CHENG**
ALLEN KONG
BOBBY PANG
YORDON KWOK

Agenda

1. Goals
2. Pipeline
3. Technical Details and Difficulties
4. Explanation on Tools
5. Demo Time
6. Q&A

GOALS

Use python to scrap tweets with keyword: #ANYWORD

- Collect tweets from ANY time interval
- Make analysis on tweet data

Pipeline

Communicate and consider our users requirement

Scrape tweets base on the time period

Store the tweet data into SQL Database

Make analysis on tweet data

Technical Details and Difficulties

Historical Data
v.s.
Filtered Stream

Selection on
Sentiment
Analysis Tool

Limitation on
Scraping
Million of
data
...
Airflow
(Developing)

Automatic
Time
Selection

Tool for scraping- snscrape

1

3rd Party Scraper

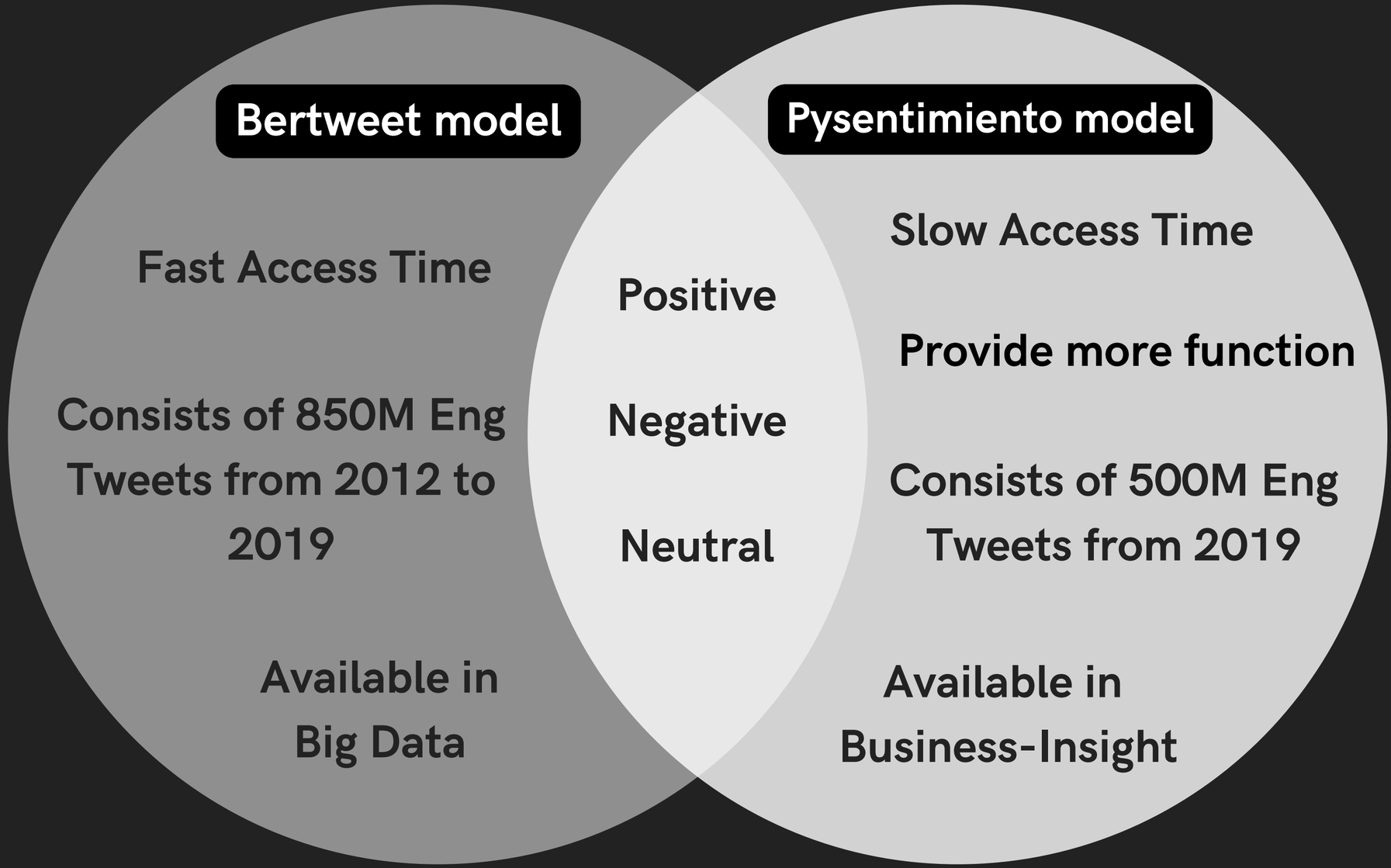
2

Unlimited Volume

3

Self-define Time Interval

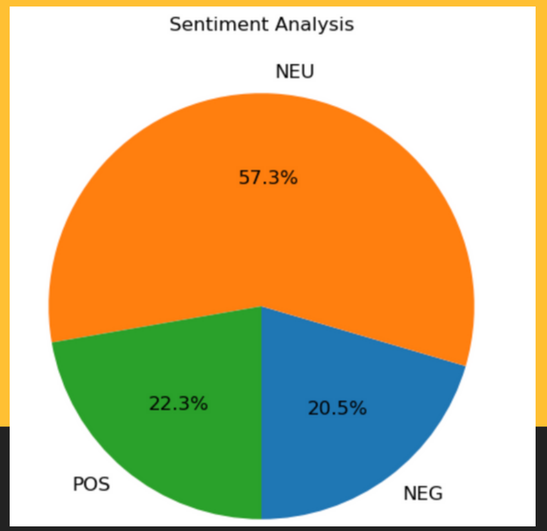
Tool for Sentiment Analysis - Transformer



Tool for plotting graph- matplotlib

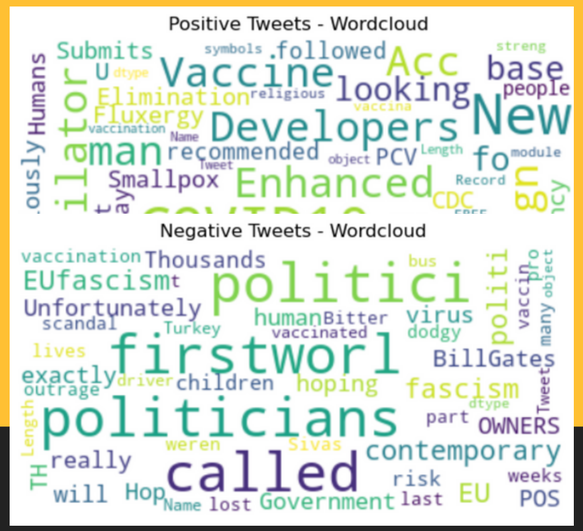
1

Piechart



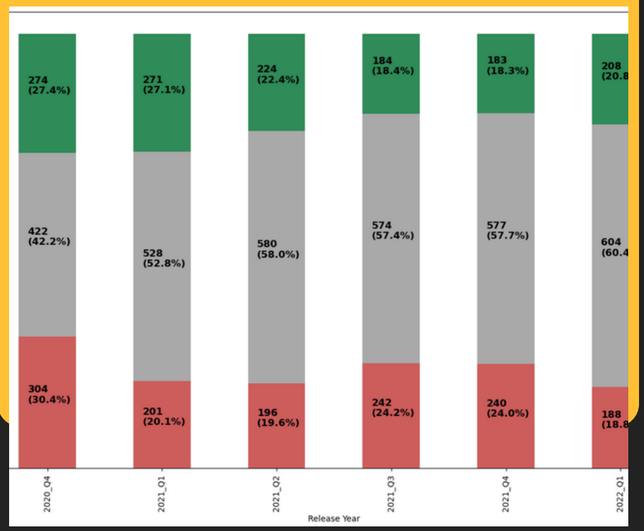
2

WordCloud



3

100% stacked column chart



Demo

Credit to

MIT License

Copyright (c) 2020-2021 VinAI Research

Permission is hereby granted, free of charge, to any person obtaining a copy
of this software and associated documentation files (the "Software"), to deal
in the Software without restriction, including without limitation the rights
to use, copy, modify, merge, publish, distribute, sublicense, and/or sell
copies of the Software, and to permit persons to whom the Software is
furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all
copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE
SOFTWARE.

- Janny
  - Random day selection
  - Auto-create schema
- Marco
  - SQL Database Engine

@inproceedings{perez2022robertuito,
  author   = {Pérez, Juan Manuel and Furman, Damián Ariel and Alonso Alemany, Laura and Luque, Franco M.},
  title   = {RoBERTuito: a pre-trained language model for social media text in Spanish},
  booktitle = {Proceedings of the Language Resources and Evaluation Conference},
  month   = {June},
  year   = {2022},
  address   = {Marseille, France},
  publisher = {European Language Resources Association},
  pages   = {7235--7243},
  url   = {https://aclanthology.org/2022.lrec-1.785}}

TIME TO
Q&A