



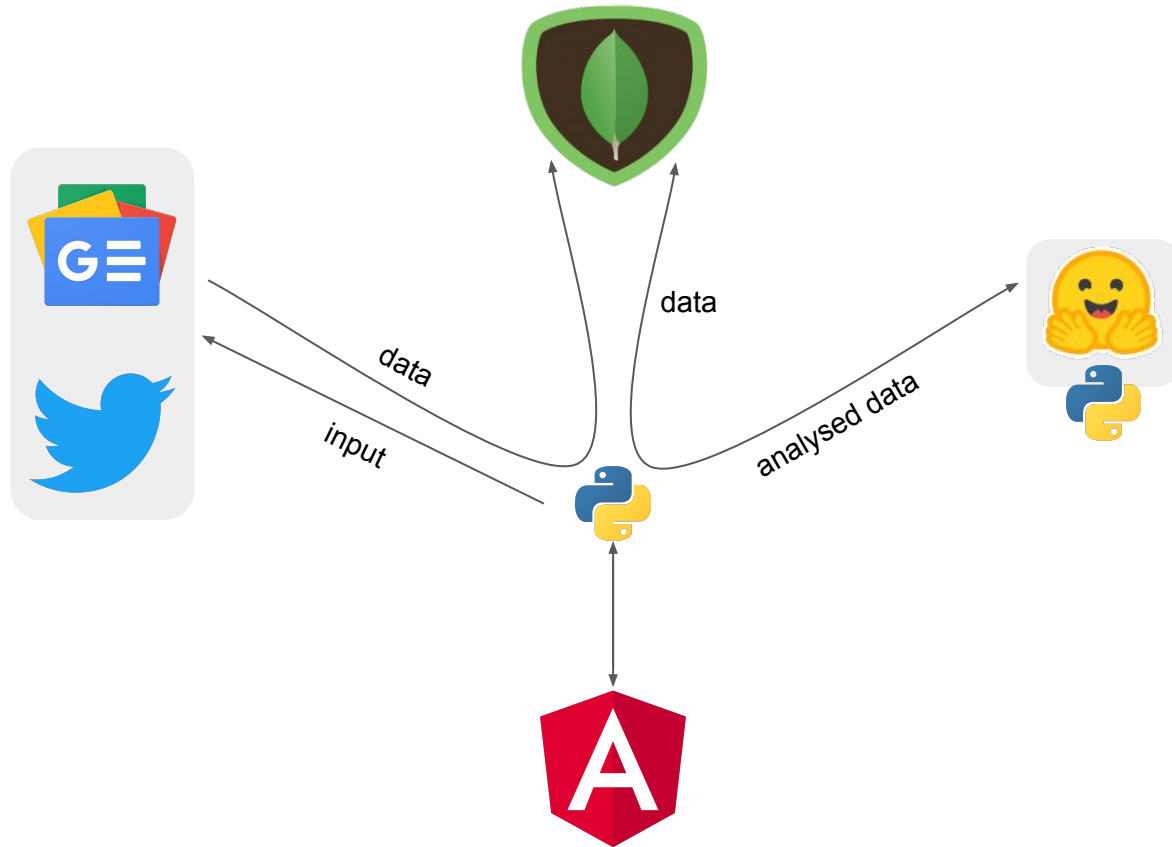
# TweetSense

## Sentiment Analysis

**Data Engineering:** Marjory QUENE, Rania RAKEZ,  
**Data Science:** Tristan PREVOT, Hugo PREVOTEAU,  
**Web Development:** Zinedine NAIT HAMMOU, Florent POUX

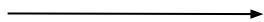
# Workflow

---



# Scrapping

---



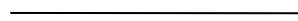
Keys twitter  
developer



Tweepy



GoogleNews

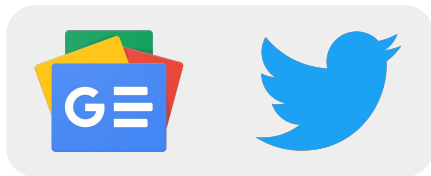


Beautiful Soup



# MongoDB Atlas

---



pymongo



- 500MB
- scalable cluster sizes

# MongoDB Atlas

twitterdb.twitter\_search

DOCUMENTS 481

Documents

Aggregations

Schema

Explain Plan

Indexes

FILTER

ADD DATA



VIEW



Display

```
{
  "id": "ObjectId(\"5fafde09b53ac84ac1cff709\")",
  "created_at": "Sat Nov 14 13:39:16 +0000 2020",
  "id": "1327607028575944704",
  "id_str": "1327607028575944704",
  "text": "RT @eraser: International Journal of Emerging Technologies in Learning...",
  "source": "<a href='\"https://mobile.twitter.com\"' rel='\"nofollow\"'>Twitter Web App</a...\"",
  "truncated": false,
  "in_reply_to_status_id": null,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null,
  "in_reply_to_user_id_str": null,
  "in_reply_to_screen_name": null,
  "user": {
    "geo": null,
    "coordinates": null,
    "place": null,
    "contributors": null,
    "retweeted_status": {
      "is_quote_status": false,
      "quote_count": 0,
      "reply_count": 0,
      "retweet_count": 0,
      "favorite_count": 0,
      "entities": {
        "favorited": false
      }
    }
  }
}
```

SHOW 5 MORE FIELDS

twitterdb.google\_search

DOCUMENTS 84

Documents

Aggregations

Schema

Explain Plan

Indexes

FILTER

ADD DATA



VIEW



Display

```
{
  "id": "ObjectId(\"5fc2afa862689634aea7e429\")",
  "created_at": "Wed November 25 11:10:48 +0000 2020",
  "date": "3 days ago",
  "datetime": "2020-11-25 11:10:48",
  "desc": "ABC News spoke with a variety of health experts to determine the top f...",
  "img": "https://lh6.googleusercontent.com/proxy/y2AfxE5KwS1bmTasn1FB8AhpNLSXMF...",
  "link": "news.google.com/.articles/CBMiaGh0dHBz0i8vYwJjbmV3cy5nby5jb20vSGVhbHR...",
  "media": null,
  "site": "ABC News",
  "text": "From wildfires to disease, here are the top 5 ways climate change is a...",
  "title": "From wildfires to disease, here are the top 5 ways climate change is a..."
}
```

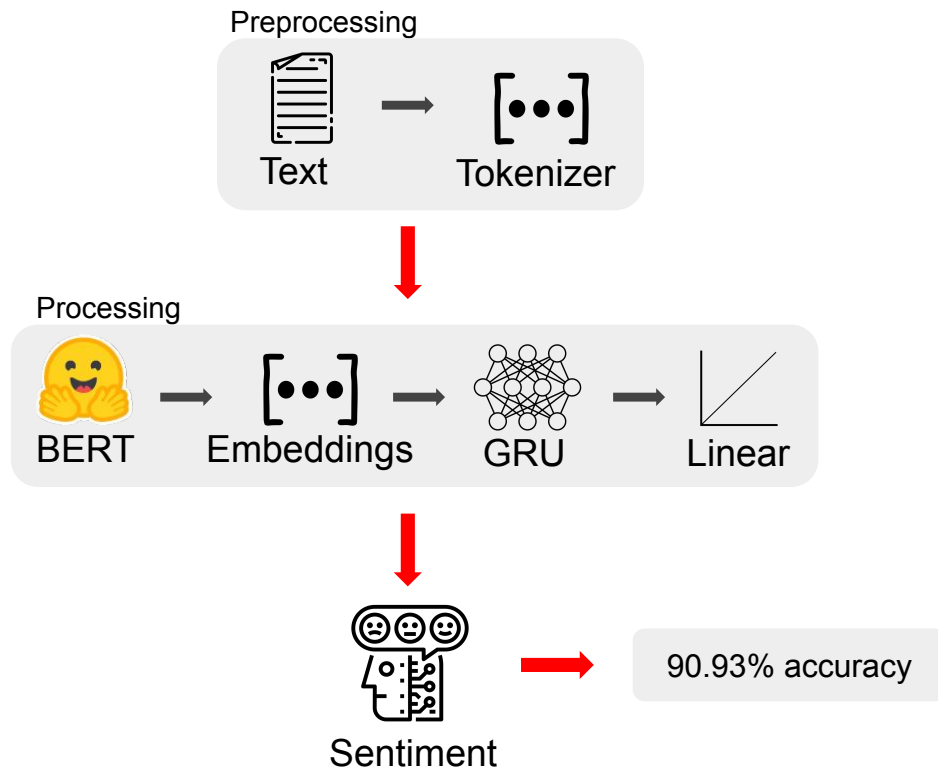
# Code

scrappingTweetsAndGoogleNewsWithConfig.py

Config.json

## BERT: fine tuning

---



CODE



# Angular 10



Open-source



Maintained



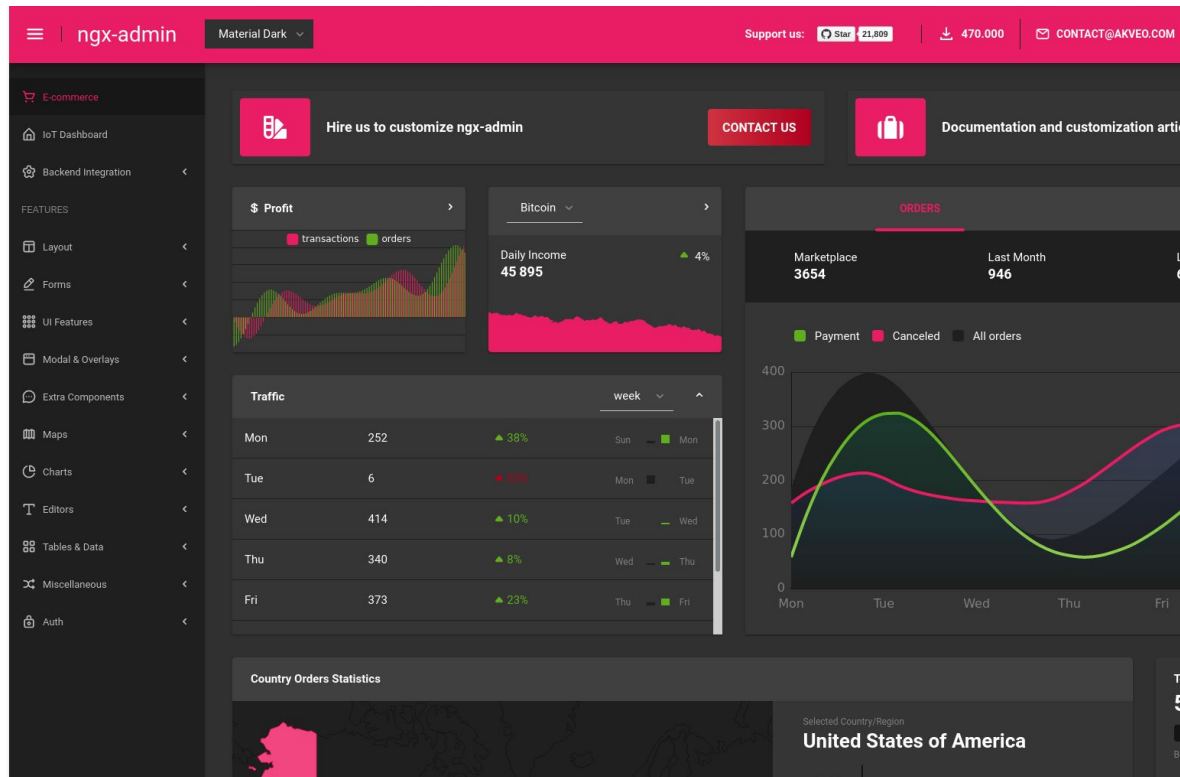
Stable



Efficient



Documented



# Production

## Security Report Summary



Site:	<a href="https://tweetsense.flowp.eu/">https://tweetsense.flowp.eu/</a>
IP Address:	134.209.236.0
Report Time:	04 Dec 2020 15:56:06 UTC
Headers:	<div>✓ Content-Security-Policy ✓ X-Frame-Options ✓ Strict-Transport-Security ✓ X-Content-Type-Options ✓ Referrer-Policy ✓ Permissions-Policy</div>
Warning:	Grade capped at A, please see warnings below.

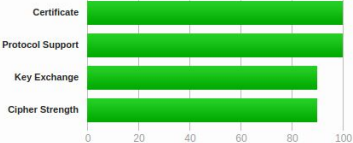
### SSL Report: tweetsense.flowp.eu (134.209.236.0)

Assessed on: Fri, 04 Dec 2020 16:03:18 UTC | [Hide](#) | [Clear cache](#)

[Scan Another »](#)

#### Summary

Overall Rating



Visit our [documentation page](#) for more information, configuration guides, and books. Known issues are documented [here](#).

This server supports TLS 1.3.

HTTP Strict Transport Security (HSTS) with long duration deployed on this server. [MORE INFO »](#)

### Found 2 open ports (1 host)

134.209.236.0

> tweetsense.flowp.eu

Port Number	State	Service Name
80	open	<a href="#">http</a>
443	open	<a href="#">https</a>



# Demonstration

## State of the Art: databases

---



# MongoDB Atlas: pricing

---

## M0

Shared RAM

512 MB storage

\$0 per month



## M2

Shared RAM

2 GB storage

\$9 per month



## M5

Shared RAM

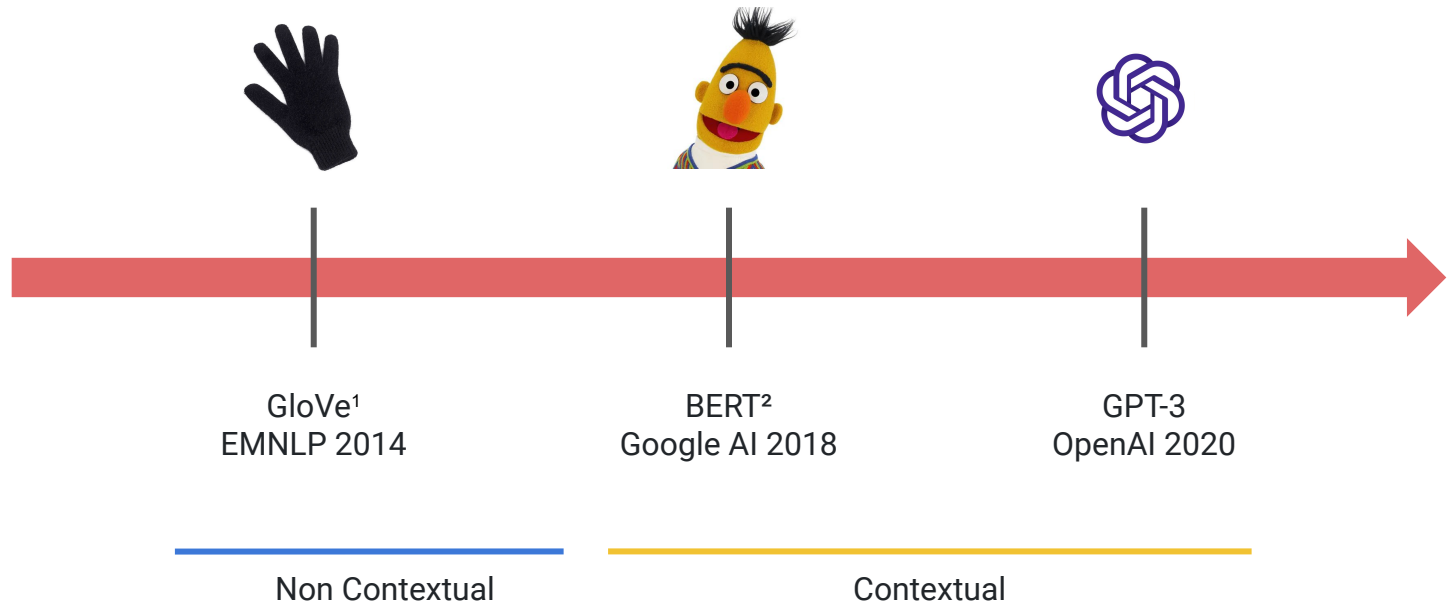
5 GB storage

\$25 per month



# State of the Art: NLP

---

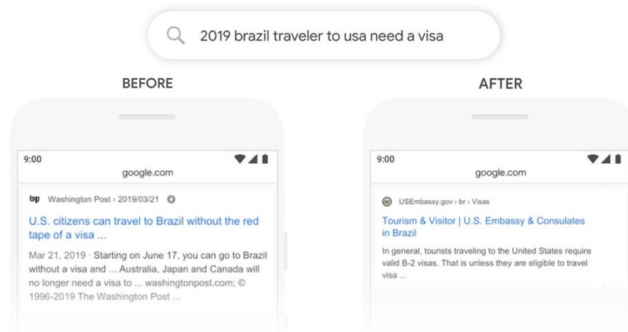
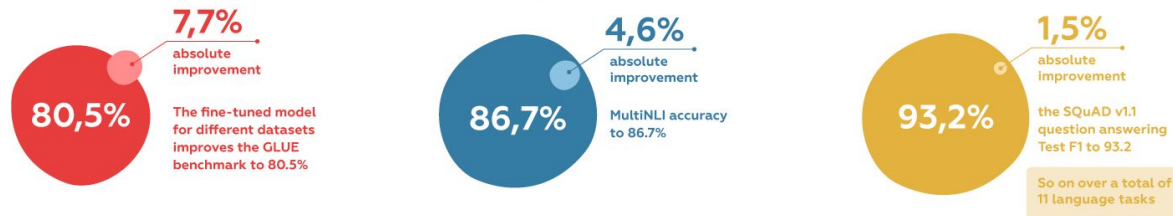


---

<sup>1</sup> Jeffrey Pennington et al.: “GloVe: Global Vectors for Word Representation”







<sup>2</sup> Jacob Devlin et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

# BERT: pros



## IMDB Benchmark









---

		Accuracy	Open Source
	GloVe	72.0%	
	BERT	90.93%	
	GPT-3	95.7%	



## GPT-3: pricing

---

Explore	Create	Build	Scale
Free	\$100/mo	\$400/mo	Contact Us
<b>100K</b> tokens Or <b>3-month trial</b> <small>Whichever comes first</small>	<b>2M</b> tokens/mo 8 cents per additional 1k tokens	<b>10M</b> tokens/mo 6 cents per additional 1k tokens	Connect with our team for access to pricing that fits your needs: <a href="mailto:support@openai.com">support@openai.com</a>
 API Playground  Developer Slack	 API Playground  Developer Slack  Priority Support	 API Playground  Developer Slack  Priority Support	



Thank you for your attention

Questions ?

## Architecture

---

