

# Covid Vaccine Sentiment Analysis

Jose Carlos Diaz Munoz  
Emanuele Travanti  
Carlos Tezna

# Problem

No existe estudios referente a la percepción del sentimiento de las distintas vacunas y los posibles razones o tópicos relacionados dichos sentimientos.

# Solution

Se busca crear un modelo de análisis de sentimientos para investigar a fondo la percepción pública de distintas vacunas a través de Twitter.

# Project Management Methodology - CRISP DM





Article

## Building a Twitter Sentiment Analysis System with Recurrent Neural Networks

Sergiu Comin Nistor<sup>1,2</sup>, Mircea Moca<sup>3</sup>, Darie Moldovan<sup>3,4</sup> and Delia Beatrice Oprean<sup>4</sup>  
and Razvan Liviu Nistor<sup>5</sup>



<sup>1</sup> Synergy Crossis CG, 10141 Tallinn, Estonia; sergiu.nistor@synergycrossis.ee (S.C.N.)

<sup>2</sup> Department of Electrical Engineering, Babes-Bolyai University, 400084 Cluj-Napoca, Romania

<sup>3</sup> Business Information Systems Department, Interdisciplinary Centre for Data Science, Babes-Bolyai University, 400084 Cluj-Napoca, Romania

<sup>4</sup> Coaching Consult, 40019 Cluj-Napoca, Romania; deliabeatrice@gmail.com

<sup>5</sup> Department of Computer Science, Babes-Bolyai University, 400084 Cluj-Napoca, Romania;

nistor@cs.ubbcluj.ro

**Abstract:** This paper presents a sentiment analysis solution on tweets using Recurrent Neural Network (RNN). The method can quantify tweets with an 86.7% accuracy rate, considering a binary task, after experimenting with 20 different design approaches. The solution integrates an attention mechanism aiming to enhance the network, with a two-way localization system at memory cells. The proposed solution is based on a hybrid architecture, which consists of a recurrent neural network and the building blocks that grounded the design decisions of our solution, employed as a core classification component within a sentiment indicator of the SynergyCrossis platform.

**Keywords:** sentiment analysis; recurrent neural network; twitter; classification; attention mechanism



Citation: Nistor, S.-C.; Moca, M.; Moldovan, D.; Oprean, D.B.; Nistor, R.L. Building a Twitter Sentiment Analysis System with Recurrent Neural Networks. *Sensors* **2021**, *21*, 2280. <https://doi.org/10.3390/s21022280>

Academic Editor: Mirjana Šukalo

Received: 10 February 2021

Accepted: 21 March 2021

Published: 24 March 2021

**Abstract:** This paper presents a sentiment analysis solution on tweets using Recurrent Neural Network (RNN). The method can quantify tweets with an 86.7% accuracy rate, considering a binary task, after experimenting with 20 different design approaches. The solution integrates an attention mechanism aiming to enhance the network, with a two-way localization system at memory cells. The proposed solution is based on a hybrid architecture, which consists of a recurrent neural network and the building blocks that grounded the design decisions of our solution, employed as a core classification component within a sentiment indicator of the SynergyCrossis platform.

**Keywords:** sentiment analysis; recurrent neural network; twitter; classification; attention mechanism

**1. Introduction**

Nowadays, social media touches all domains of the social activity. People share information in order to better understand and inform others, related to things they care for, e.g., products, services, events, or places. A global system like Twitter allows people to express their feelings in a relatively simple message.

Understanding what is related to products/services is important both for the decision-makers that control the respective products/services and also for their consumers. Building aggregate knowledge for decision-makers can be done in the form of documents, reports, and dashboards, but also through sentiment analysis solutions.

Building indicators for measuring the sentiment of people on a specific topic of discussion is a very hard task. Among others, it requires building a performance natural language processing (NLP) model that can extract the sentiment from the sequence of characters with high accuracy.

There are various models for extracting sentiment from text and more specifically from tweets. However, building a sentiment indicator for a particular set of topics of interest requires building a complete solution that is designed and validated for the specific needs.

Recurrent neural networks (RNNs) are a variation of artificial neural networks that can model sequences [1]. The problem of sentiment analysis can be viewed as a sequence classification problem. The input is the text which is a sequence of tokens. These tokens can be, for example, words or characters. The output may be a single value indicating the sentiment expressed.

Modern RNNs are based on memory cells, which are also to learn dependencies between data points placed further apart in the sequence. Attention mechanisms [2]

## Public Sentiment Analysis of COVID19 Vaccination Drive in India

Dr. Akash D Dubey

Jaipuria Institute of Management, Jaipur

**Abstract:** The vaccination drive for prevention against COVID19 in India started from 16<sup>th</sup> January 2021. Two vaccines used for this drive are Oxford-AstraZeneca's Covishield and Bharat Biotech's Covaxin. This drive has already crossed 600,000 marks in first four days and the government has stated that the drive will be completed in 150 days to ensure maximum safety to the citizens of the country. However, there is a section of the population which is still negative towards the COVID19 vaccination. This research work has been conducted to analyze the sentiments in the tweets posted in India regarding these two vaccines. The analysis shows that while a majority of the population is posting with positive sentiments towards these vaccines, there are also negative sentiments associated with them, associated with the emotions such as fear and anger.

**Keywords:** COVID19, Vaccine, India, Covaxin, Covishield

### Introduction

COVID-19, or commonly known as COVID-19 has affected more than 96 million population and two million deaths throughout the world [1]. While majority of the world has faced complete or partial lockdown in the year 2020, it has become economically infeasible to continue with it. In the current scenario, the necessity of vaccination has become eminent. India, being the second most affected country from COVID19 announced its vaccination drive from 16<sup>th</sup> January 2021, after Central Drugs and Standards Committee (CDSC) formally approved vaccines by Bharat Biotech and Serum Institute of India (SII)[2].

The first phase of the vaccination drive aims to involve 30 million healthcare professionals and 270 million priority beneficiaries and is expected to end by July (3). The timeline for the rest of the population has not been specified yet. Moreover, it has been reported that around 92 countries have approached India for the vaccination, especially after the reports of dramatic decrease in the efficacy of Chinese Vaccination Sinovac (4).

Since the Indian government is proceeding with two vaccines for the drive, namely Covishield and Covaxin, the Indian citizens has already announced this due to high demand and logistics challenges. The individuals will be able to choose the vaccines and will have to proceed per the government plans (5). While the majority of the Indian population awaits their chance to get vaccinated, they are also sharing their thoughts on Twitter. This research study aims at analyzing the sentiments and emotions of the people of India regarding both the vaccines i.e. Covishield and Covaxin.

## An NLP and LSTM Based Stock Prediction and Recommender System for KOSDAQ and KOSPI



Master of Science in Computer Science  
June 2019

Indranath Chatterjee<sup>1</sup>, Jeon Gwan<sup>2</sup>, Yong Jin Kim<sup>2</sup>, Min Seok Lee<sup>2</sup>,  
and Miyoung Cho<sup>2,✉</sup>

<sup>1</sup> Department of Computer Engineering, Tongmyong University,  
Busan 48520, South Korea

<sup>2</sup> Department of Game Engineering, Tongmyong University,  
Busan 48520, South Korea  
<http://gch@tongyong.ac.kr>

**Abstract:** Stock market prediction is one of the complex analysis of all time. Different expert analysts, as well as computer scientists, are working on the prediction of a stock price. One of the major challenges is the prediction of future stock values. The primary challenge is the nature of the movement of the daily price, which depends on various factors. To build a predictive model for the analysis of stock data and prediction is an active area of research. However, we found only a few numbers of studies performed on the prediction of stock prices. In this paper, we propose a model that can predict the future value of a stock. This study proposed a three-stage approach based on Natural Language Processing and Deep Learning techniques to analyze, compute the sentiment of the tweets posted on Twitter and predict the future value of a stock. This study involves the application of natural processing techniques and deep learning techniques around 2500 Korean companies listed on KOSDAQ and KOSPI. Finally, this paper not only fully presents the current condition of the stock and overall Korean stock exchange; secondly, it recommends the potential months and weeks for investment, and finally, it predicts the future value of a stock with high accuracy. This paper may pose as a structural framework for developing a complete stock market prediction application.

**Keywords:** Stock market prediction · Recurrent neural network · LSTM · Natural language processing · KOSDAQ · KOSPI

### 1 Introduction

A stock market, also known as the share market, is a marketplace combined with a financial market where shares of companies are traded. It is a kind of ownership claim on a company when someone buys a share of its business. Apart from liquid trading, people take advantage of it in the form of investment. The stock market trading and investment are done using either stockbrokers or online trading platforms.

© Springer Nature Switzerland AG 2021  
M. Cho et al. (Eds.), JHIC 2020, LNCS 1615, pp. 409–413, 2021.  
[https://doi.org/10.1007/978-3-030-68449-5\\_37](https://doi.org/10.1007/978-3-030-68449-5_37)

## Sentiment Analysis of Twitter Data Using Machine Learning and Deep Learning Methods

Kundan Reddy Manda



Faculty of Computing, Blekinge Institute of Technology, 371 70 Karlskrona, Sweden

Company	Tipo	Dosis Requeridas	Almacenamiento	Tiempo de vencimiento
Oxford UniAstraZeneca	Vector viral modificado	2	2-8°C	6 meses
Moderna	RNA	2	-25 a -15°C	7 Meses
Pfizer-BioNtech	RNA	2	-80 a -60°C	6 meses
Sputnik V (gamaleya)	Vector viral modificado	2	2 - 8°C	--
Sinovac	Virus inactivo	2	2 - 8°C	--
NovaVax	Protéico	2	2 - 8°C	--
Janssen	Vector viral modificado	1	2 - 8°C	3 meses

Fuente: <https://www.bbc.com/news/world-us-canada-56410179>

Pronóstico ventas (USD Bill.)					
Desarrollador	Precio (USD**)	2021	2022	2023	% variación precio de acción Mar 20 - Mar 21
BioNTech/Pfizer**	37.5	21.5	8.6	2	156
Moderna	36.5	19.6	12.2	11.4	372
Johnson & Johnson	10	10	-	-	7.7
AstraZeneca	7.2	1.9	3	-	-8.6
Sinovac	27.2	-	-	-	-21.6
Gamaleya	20	-	-	-	-
Novavax	3	-	-	-	1,128

- Planificación de estudios locales
- Elemento de toma de decisión para gestión pública
- Planificación de estrategias informativas
- Prevención de pérdida de inversión en compra de vacunas

# Datasets

Training/Test/Validation (model) Dataset: <https://www.kaggle.com/kazanova/sentiment140>

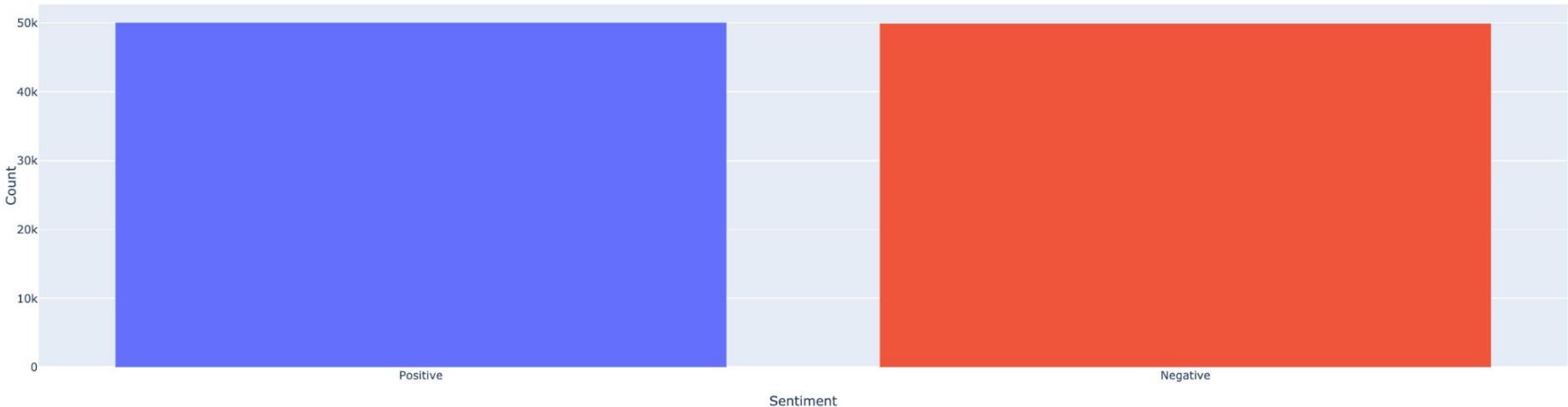
<b>sentiment</b>	<b>id</b>	<b>date</b>	<b>query</b>	<b>user</b>	<b>tweet</b>
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli

Covid Vaccine (research) Dataset:

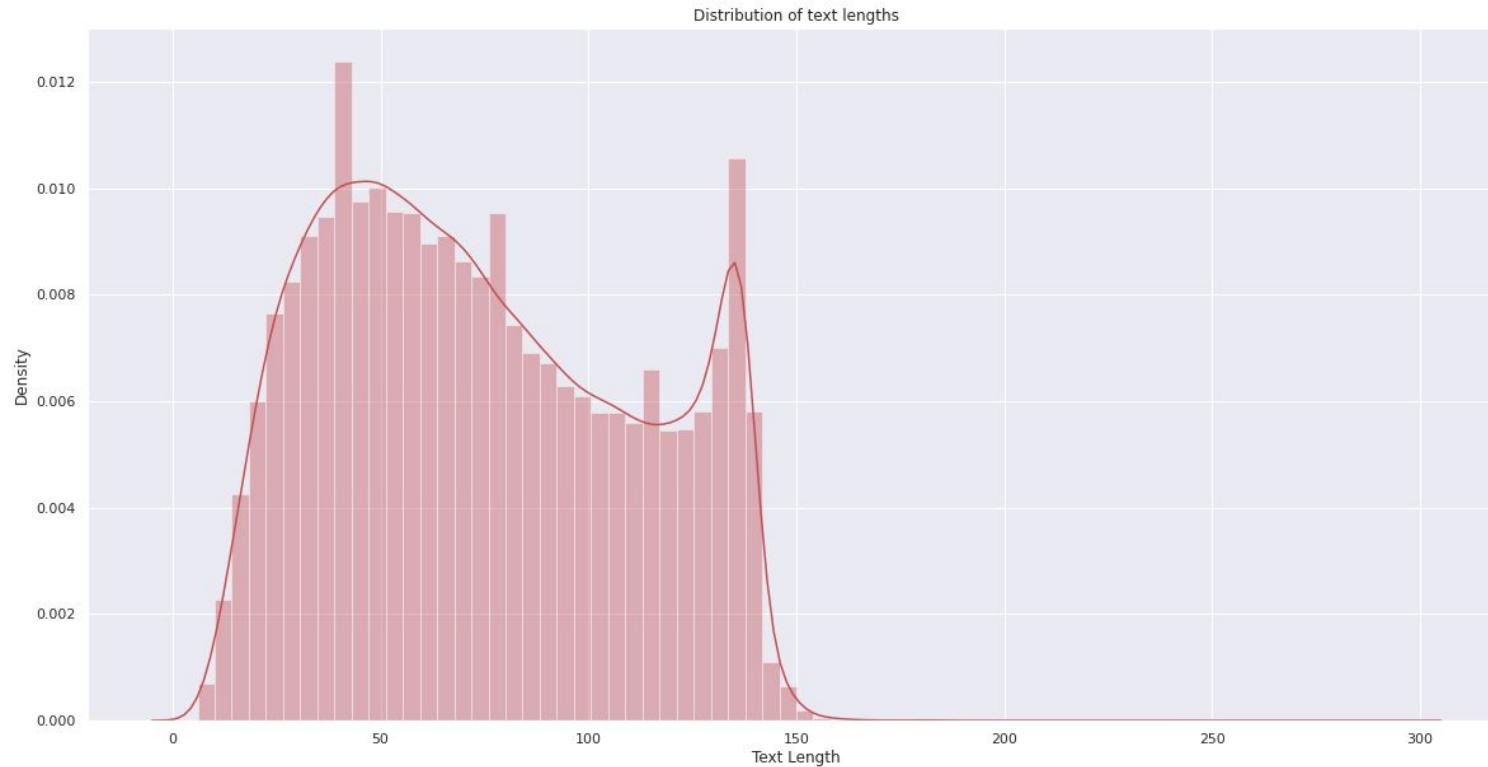
<https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets>

<b>id</b>	<b>user_name</b>	<b>user_location</b>	<b>user_description</b>	<b>user_created</b>	<b>user_followers</b>	<b>user_friends</b>	<b>user_favourites</b>	<b>user_verified</b>	<b>date</b>	<b>text</b>	<b>hashtags</b>	<b>source</b>	<b>retweets</b>	<b>favorites</b>	<b>is_retweet</b>
0	1340539111971516416	Rachel Roth	La Crescenta-Montrose, CA Aggregator of Asian American news; scanning di...	2009-04-08 17:52:46	405	1692	3247	False	2020-12-20 06:06:44	Same folks said daikon paste could treat a cyt...	[PfizerBioNTech]	Twitter for Android	0	0	False
1	1338158543359250433	Albert Fong	San Francisco, CA Marketing dude, tech geek, heavy metal & 80s ...	2009-09-21 15:27:30	834	666	178	False	2020-12-13 16:27:13	While the world has been on the wrong side of...	NaN	Twitter Web App	1	1	False
2	1337858199140118533	elli <small>🕒</small> 🕒	Your Bed	2020-06-25 23:30:28	10	88	155	False	2020-12-12 20:33:45	#coronavirus #SputnikV #AstraZeneca #PfizerBio...	[coronavirus, 'SputnikV', 'AstraZeneca', 'Pf...	Twitter for Android	0	0	False
3	133785739918835717	Charles Adler	Vancouver, BC - Canada Hosting "CharlesAdlerTonight" Global News Radi...	2008-09-10 11:28:53	49165	3933	21853	True	2020-12-12 20:23:59	Facts are immutable, Senator, even when you're...	NaN	Twitter Web App	446	2129	False
4	1337854064604966912	Citizen News Channel	NaN Citizen News Channel bringing you an alternati...	2020-04-23 17:58:42	152	580	1473	False	2020-12-12 20:17:19	Explain to me again why we need a vaccine @Bo...	[whereareallthesickpeople, PfizerBioNTech]	Twitter for iPhone	0	0	False

Sentiment Counts

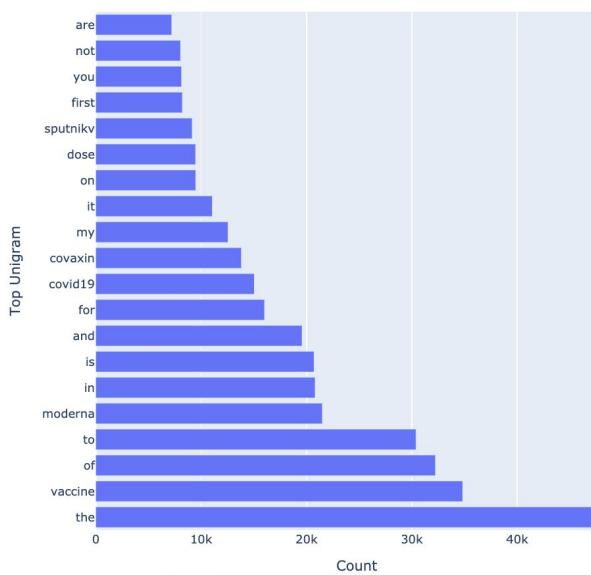


Not completely balanced, but sufficient for our problem.

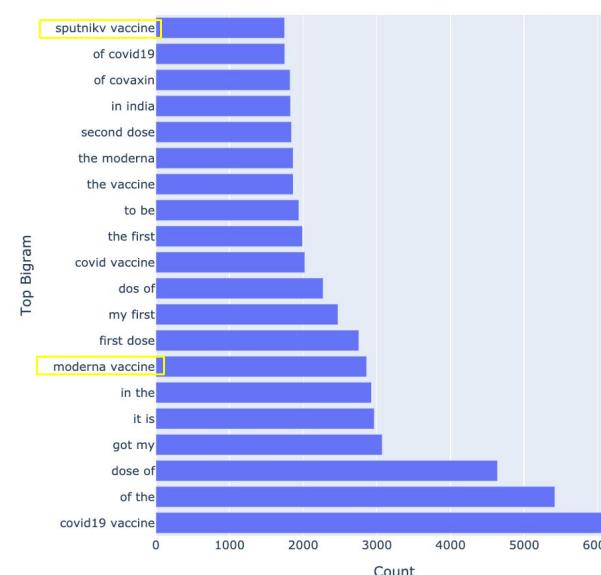




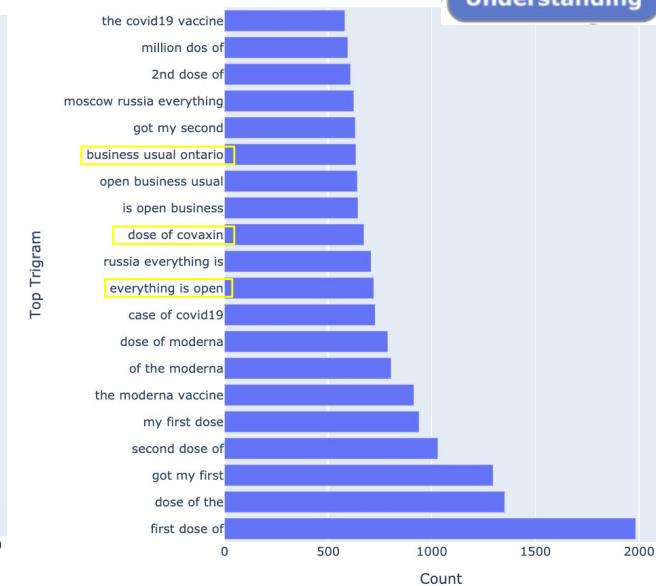
Common Unigram in Text



Common Bigram in Text



Common Trigram in Text



## Named Entity Recognition

More than 1,000 CARDINAL people in the US GPE have died of coronavirus nearly every day this month DATE

#USA # MONEY died... <https://t.co/39XmbOcFvB>

@TeraceGarnier

@NIH @Nexsy Point ORG

is us has failed again. . And

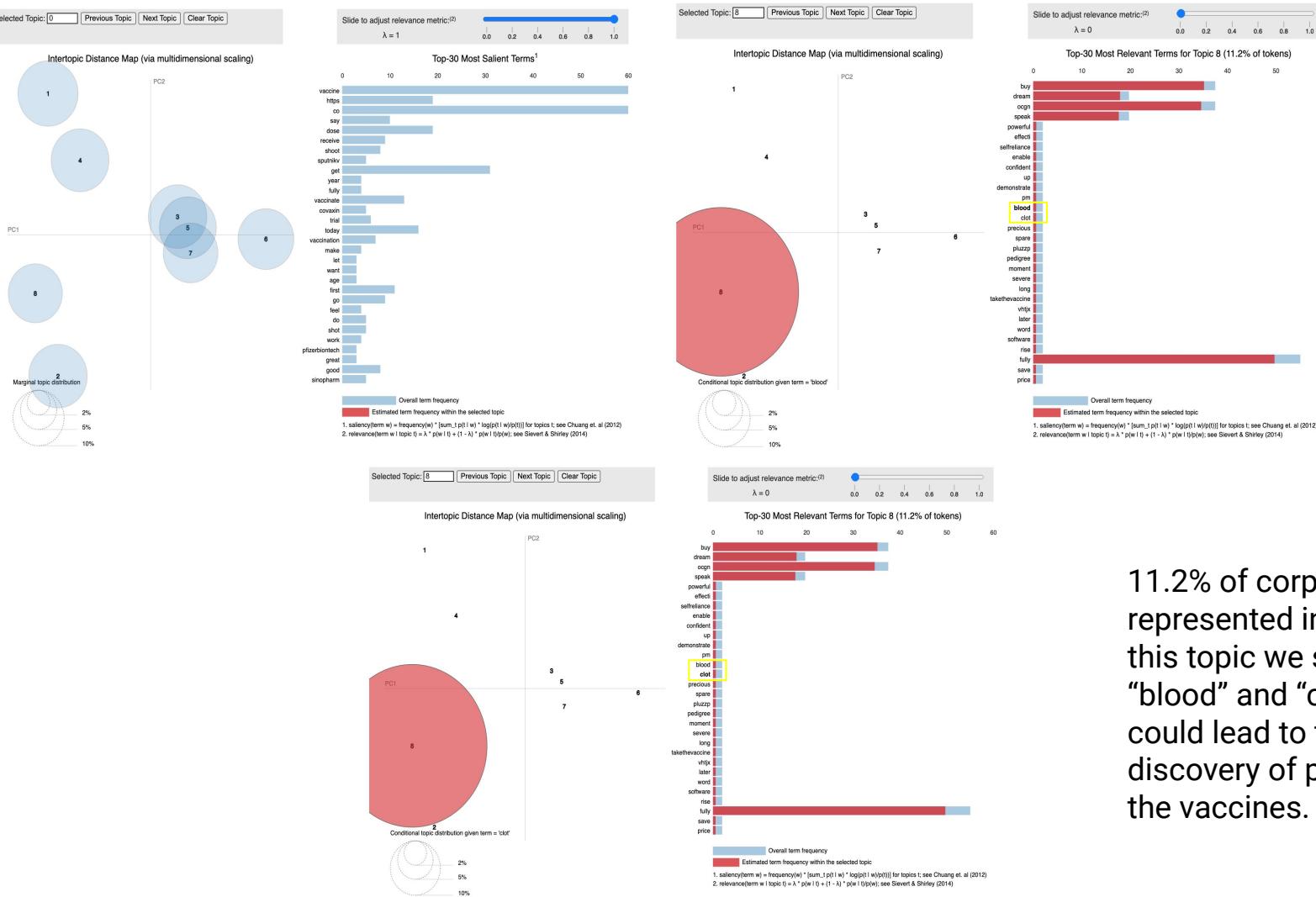
Russia GPE

has delivered. The vaccine #

CovidVaccine PERSON

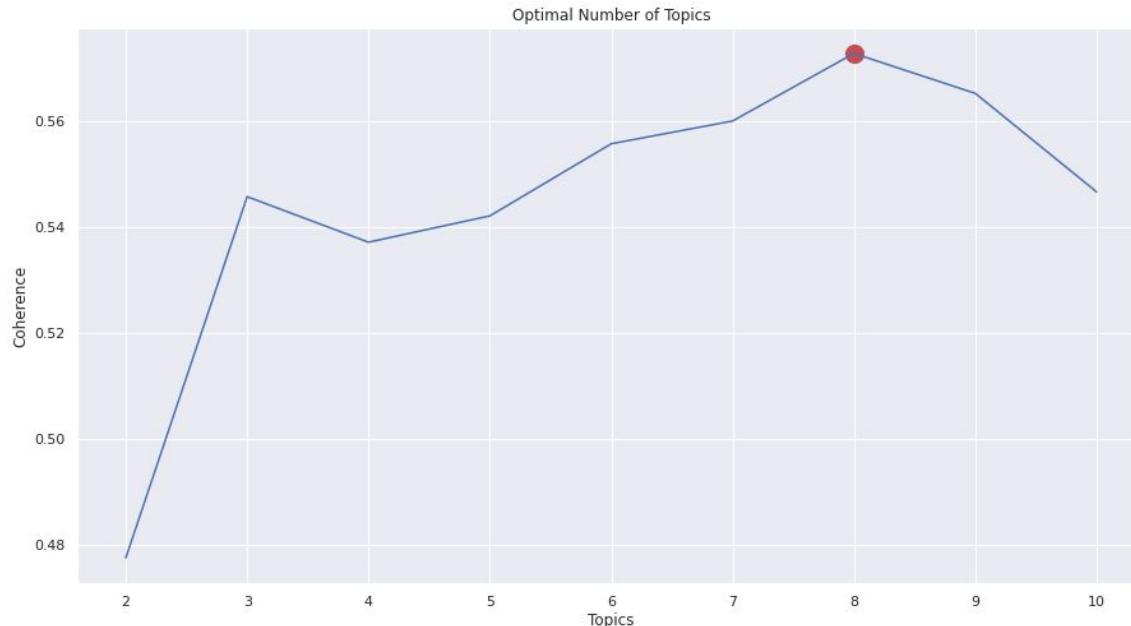
<https://t.co/RiK6XXwzHe>

## Data Understanding



11.2% of corpus is represented in topic 8. In this topic we see the terms "blood" and "clot". This could lead to the possible discovery of problems with the vaccines.

Validation_Set	Topics	Alpha	Beta	Coherence
465      100% Corpus	8	0.91	0.01	0.572675



## Grid Search Hyperparameter Tuning for Maximum Coherence Score

Hyperparameters:

- Number of Topics (K)
- Dirichlet hyperparameter
- **alpha**: Document-Topic Density
- Dirichlet hyperparameter
- **beta**: Word-Topic Density

**What is coherence score?**  
Coherence measures the relative distance between words within a topic. Syed and Spruit tested coherence scores over many datasets, and configurations, and most coherence scores were between 0.3 and 0.6.<sup>[2]</sup>

<sup>[2]</sup>Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation  
<http://www.saf21.eu/wo-content/uploads/2017/09/5004a165.pdf>

## Text Preprocessing:

- **Clean text:** remove HTML, links, symbols, etc
- **Expand contractions:** “he’ll -> he will”
- **Remove stop words:** keep negative stop words like “not”
- **Normalize text:** lowercase, apply stemming or lemmatization
- **Make n-grams:** unigrams, bigrams, and trigrams

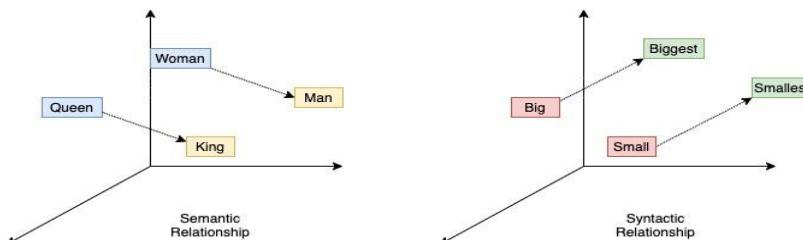
## Stemming vs Lemmatization



## This is Big Data AI Book

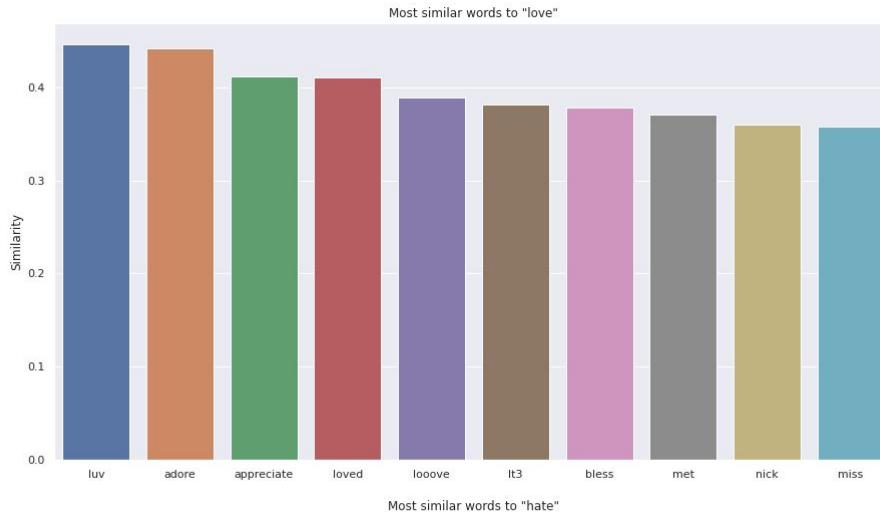
## Feature Engineering:

- **Word embeddings:** word2vec

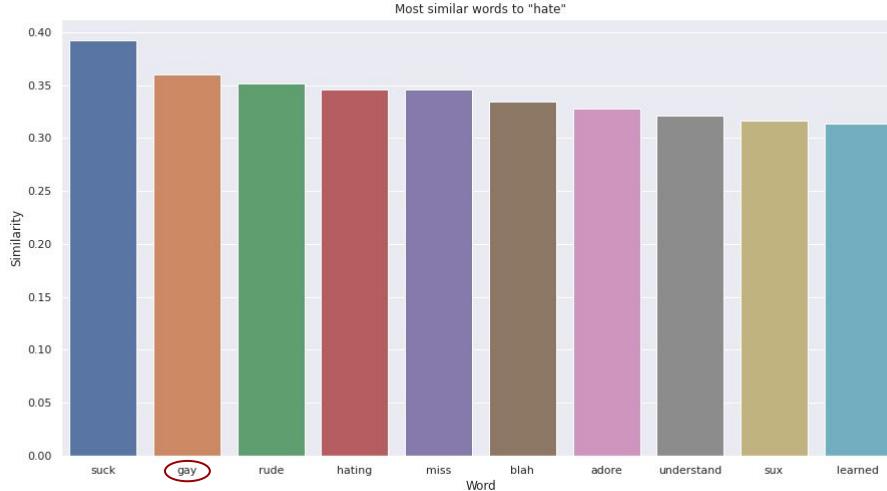


<b>Uni-Gram</b>	This	Is	Big	Data	AI	Book
<b>Bi-Gram</b>	This is	Is Big	Big Data	Data AI	AI Book	
<b>Tri-Gram</b>	This is Big	Is Big Data	Big Data AI	Data AI Book		

	text	sentiment	processed_text	processed_text_stopwords
541200	@chrishasboobs AHHH I HOPE YOUR OK!!!	0	chrishasboobs ahhh i hope your ok	chrishasboobs ahhh hope ok
750	@misstoriblack cool, i have no tweet apps fo...	0	[misstoriblack cool i have no tweet apps for my...]	[misstoriblack cool tweet apps razr 2
766711	@TiannaChaos i know just family drama. its la...	0	tiannachaos i know just family drama it lamehey...	tiannachaos know family drama lamehey next tim...
285055	School email wont open and i have geography ...	0	school email will not open and i have geography...	school email open geography stuff revise stupi...
705995	upper airways problem	0	upper airway problem	upper airway problem

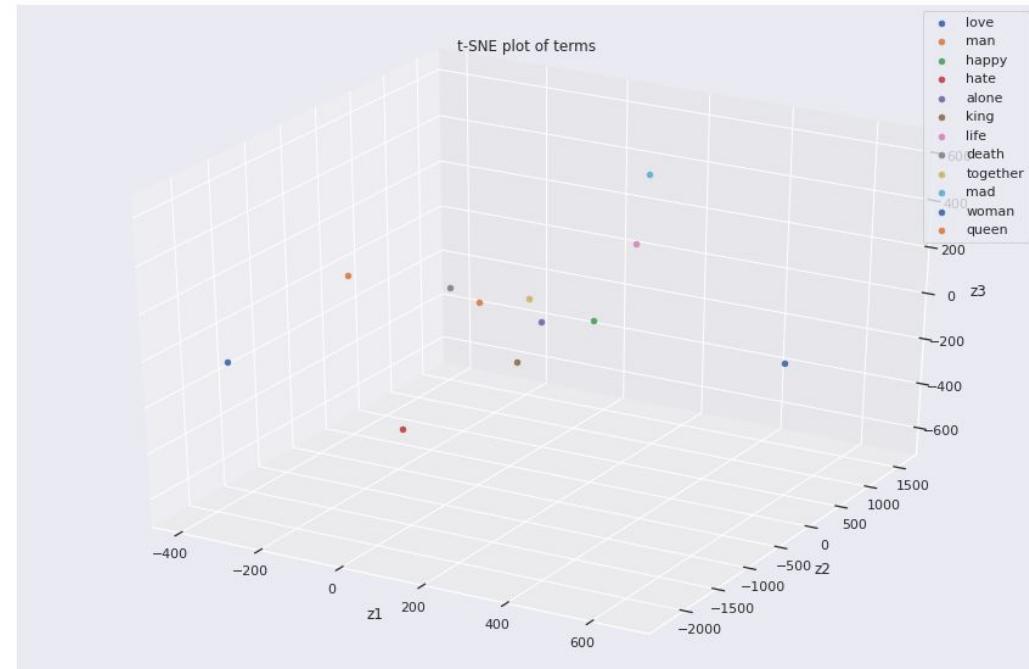
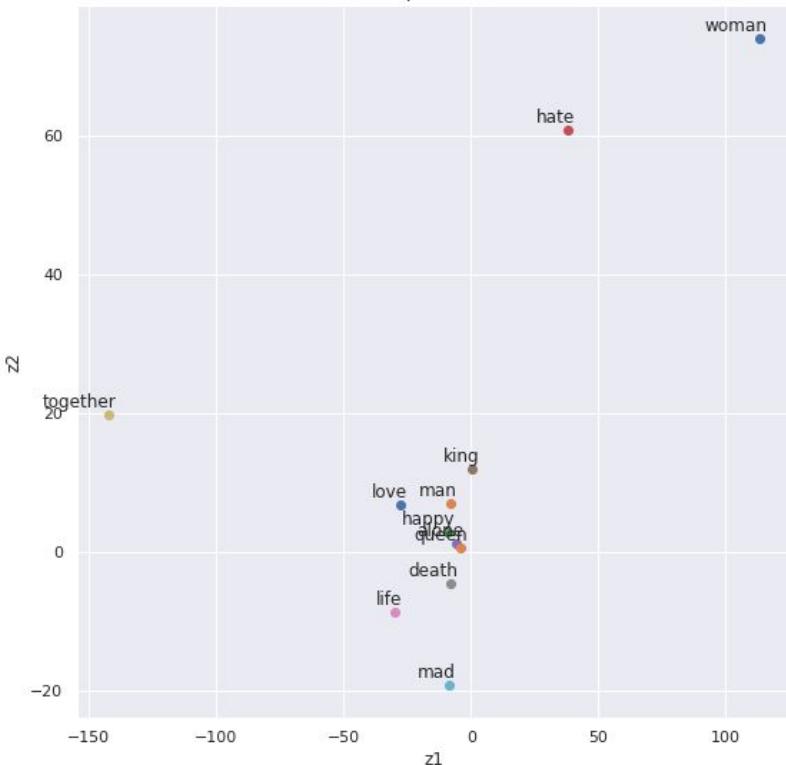


Unfortunately, we can see a connection between the words "hate" and "gay". This could show evidence of hate speech in our texts.



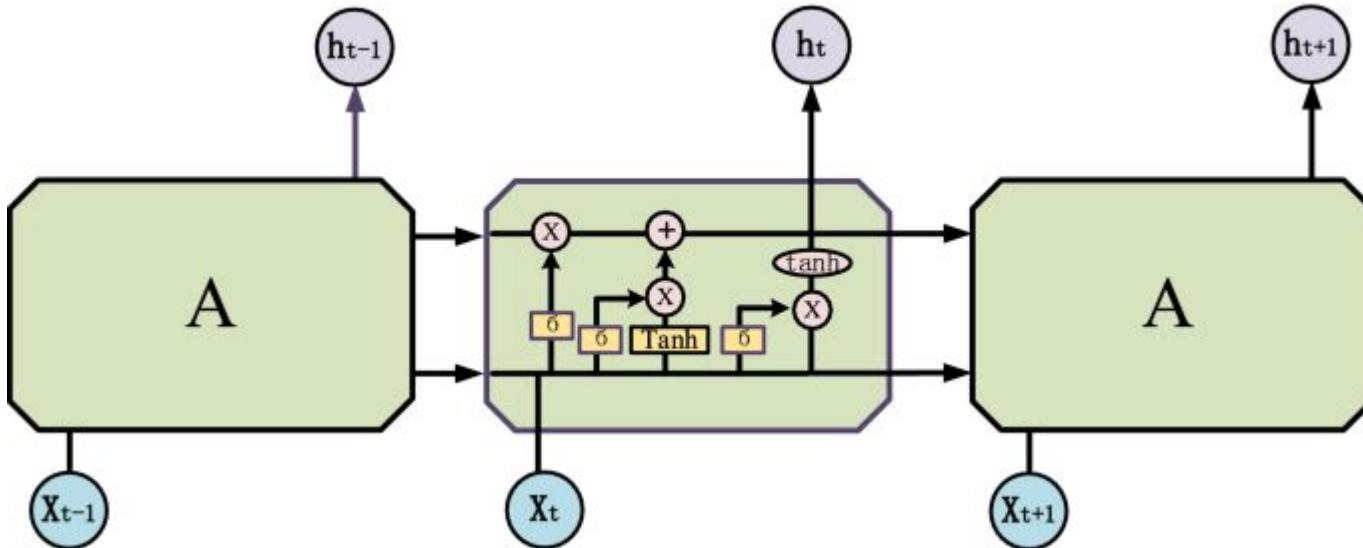
Word embeddings produced by the word2vec model allows for similarity between words to be calculated while maintaining semantic and syntactic information of each word and its use within the corpus.

t-SNE plot of terms



The t-SNE (t-distributed Stochastic Neighboring Embedding) Model is a tool to visualize high-dimensional data. It converts similarities between data points to joint probabilities and tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data. t-SNE has a cost function that is not convex, i.e. with different initializations we can get different results. Contrary to PCA it is not a mathematical technique but a probabilistic one.<sup>[1]</sup>

<sup>[1]</sup>[Visualising high-dimensional datasets using PCA and t-SNE in Python](#)



## Long Short Term Memory (LSTM) Cell

"Recurrent networks ... have an internal state that can represent context information. ... [they] keep information about past inputs for an amount of time that is not fixed a priori, but rather depends on its weights and on the input data."<sup>[1]</sup>

**Building a Twitter Sentiment Analysis System with Recurrent Neural Networks**

Sergio Coimbra Neto<sup>1,2</sup>, Mírcia Meira<sup>1</sup>, Daris Moldovan<sup>1,3</sup> and Délio Batista Oliveira<sup>4</sup>  
and Edilson Lira Nunes<sup>1,2</sup>

<sup>1</sup> Strategy Growth (SG), 04170-000, São Paulo, Brazil; <sup>2</sup> Institute of Engineering of São Paulo (IEST), University of São Paulo (USP), São Paulo, Brazil;  
<sup>3</sup> Department of Computer Science, Babes-Bolyai University, 4000 Cluj-Napoca, Romania;  
<sup>4</sup> Department of Electrical and Computer Engineering, University of São Paulo (USP), São Paulo, Brazil;

Master of Science in Computer Science  
Issue 2019



## Sentiment Analysis of Twitter Data Using Machine Learning and Deep Learning Methods

Kundan Reddy Manda

[1] Yoshua Bengio, et al., Learning Long-Term Dependencies with Gradient Descent is Difficult, 1994.

**Building a Twitter Sentiment Analysis System with Recurrent Neural Networks (RNNs)**

Sergio Coimbra Neto<sup>1,2</sup>, Mírcia Meira<sup>1</sup>, Daris Moldovan<sup>1,3</sup> and Délio Batista Oliveira<sup>4</sup>  
and Edilson Lira Nunes<sup>1,2</sup>

<sup>1</sup> Strategy Growth (SG), 04170-000, São Paulo, Brazil; <sup>2</sup> Institute of Engineering of São Paulo (IEST), University of São Paulo (USP), São Paulo, Brazil;  
<sup>3</sup> Department of Computer Science, Babes-Bolyai University, 4000 Cluj-Napoca, Romania;  
<sup>4</sup> Department of Electrical and Computer Engineering, University of São Paulo (USP), São Paulo, Brazil;

**Abstract:** This paper presents a sentiment analysis solution on tweets using Recurrent Neural Networks (RNNs). The method is classifying tweets as an RNNs accuracy rate, considering a long-term dependency between words. The proposed system uses an attention mechanism to attend one feature among all features in the network, as well as a very low loss rate of errors, calculating the average error rate of 0.0001. The system is able to classify the sentiment of the tweet and the building blocks that provided the decision of one sentence, employed as a core component of the system. The system is able to classify the sentiment of the tweet and the building blocks that provided the decision of one sentence, employed as a core component of the system.

**Keywords:** sentiment analysis, recurrent neural networks, neural classification, attention mechanism

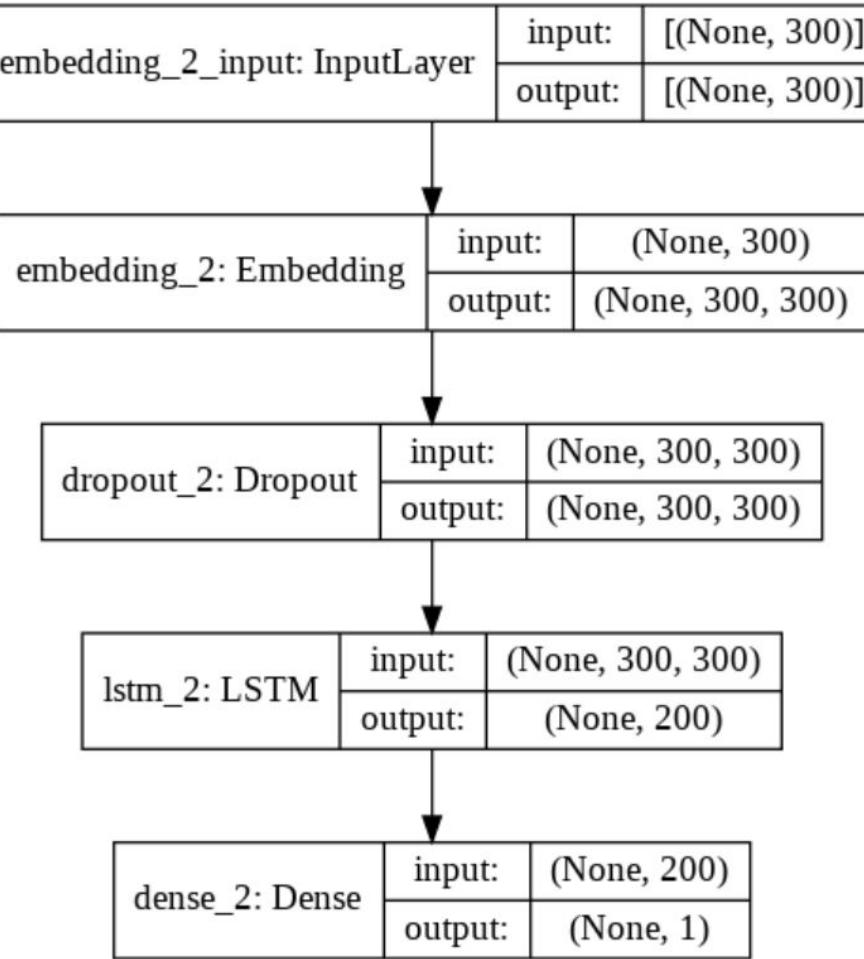
**1. Introduction**  
Nowadays, social media reaches all corners of the world, acting as people share information, write, read, like, comment, and receive others, related to things they care for, e.g., products, services, events, or places. A global system like Twitter allows people to express their opinions and feelings about the things they care for.

**2. Problem Statement**  
Understanding what people feel related to products or services is important both for companies and individuals. Companies can use this information to better serve their consumers. Building aggregate knowledge for decision-making can be done in the form of sentiment analysis, which is a process of extracting the sentiment of a text.

**3. Solution**  
Building solutions for measuring the sentiment of people on a specific topic of interest is a challenging task. In this paper, we propose a solution based on a natural language processing (NLP) solution that is capable of extracting sentiment from sequences of words. The literature presents various models for extracting sentiment from text and more specifically from Twitter. However, the extraction of sentiment from a sequence of tokens requires building a complex solution that is designed and validated for this specific problem.

**4. Methodology**  
Recurrent neural networks (RNNs) are a variation of artificial neural networks that can learn temporal dependencies between inputs and outputs, making them suitable for time series problems. The input is the text which is a sequence of tokens. These tokens can be words, characters, or even symbols. The output is the sentiment of the text, which is the sentiment expressed.

**5. Results**  
The results show that the proposed system is able to classify the sentiment of the tweet and the building blocks that provided the decision of one sentence, employed as a core component of the system.



Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 300, 300)	12589500
dropout_2 (Dropout)	(None, 300, 300)	0
lstm_2 (LSTM)	(None, 200)	400800
dense_2 (Dense)	(None, 1)	201

Total params: 12,990,501

Trainable params: 401,001

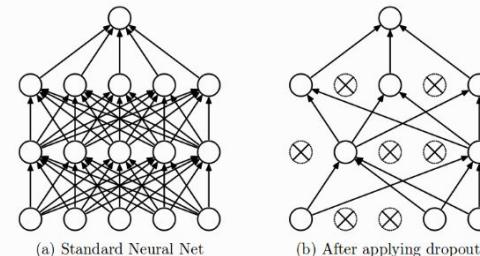
Non-trainable params: 12,589,500

Embedding layer is used to receive integer encoded text data. We are happy with the results of the learned embeddings done by the *word2vec* model so this layer will not be trainable. Only the LSTM and Dense layers will be trained, a type of transfer learning.

The Dropout layer randomly sets input units to 0 with a frequency of a predefined rate (0.5) at each step during training time, which helps prevent overfitting. Note that the Dropout layer only applies when training, such that no values are dropped during inference.

The LSTM layer has 200 units and also uses a dropout of 0.2.

The last layer (Dense) acts as a logistic regression, using a sigmoid function, and outputs the final inference.



# Compile and Train Model

```
model.compile(loss='binary_crossentropy', optimizer="adam", metrics=[ 'accuracy'])

callbacks = [ ReduceLROnPlateau(monitor='val_loss', patience=5, cooldown=0),
              EarlyStopping(monitor='val_accuracy', min_delta=1e-4, patience=5) ]

history = model.fit(x_train, y_train,
                     batch_size=BATCH_SIZE,
                     epochs=32,
                     validation_split=0.1,
                     verbose=1,
                     callbacks=callbacks)
```

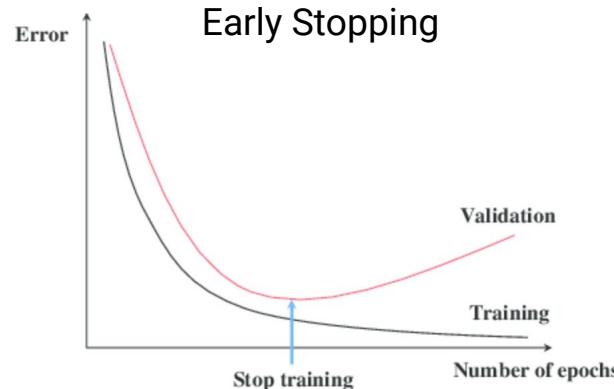
During training, 10% of the training data is set aside to be used as validation data during the current epoch.

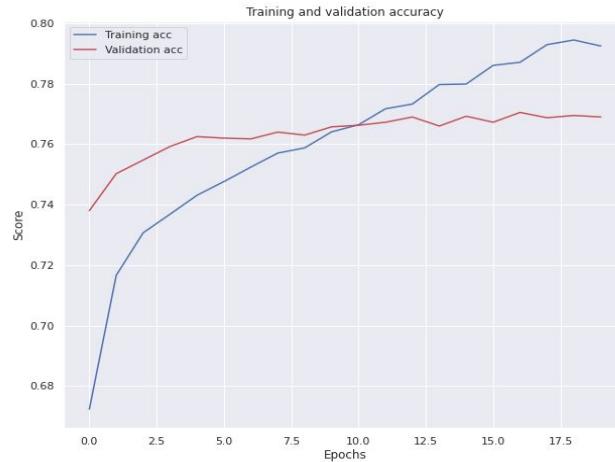
## Callbacks

- [Reduce Learning Rate on Plateau \(ReduceLROnPlateau\)](#):  
Reduce learning rate when a metric has stopped improving. Models often benefit from reducing the learning rate by a factor of 2-10 once learning stagnates. This callback monitors a quantity(validation loss) and if no improvement is seen for a 'patience' number of epochs, the learning rate is reduced.
- [Early Stopping](#):  
Stop training when a monitored metric, in this case validation accuracy, has stopped improving. Uses 'patience' parameter as well.

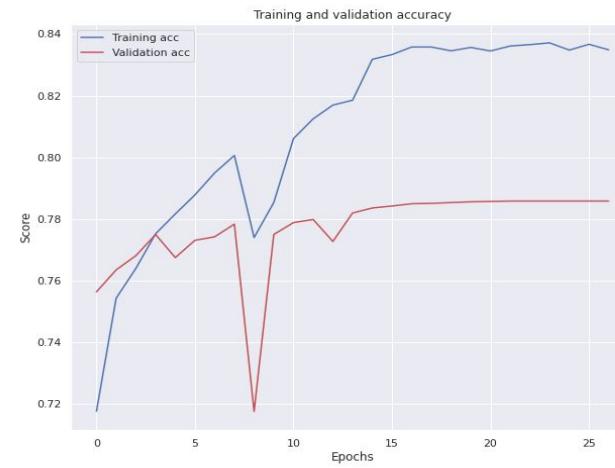
Adam optimization is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments. According to Kingma et al., 2014, the method is "computationally efficient, has little memory requirement, invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data/parameters".<sup>[1]</sup>

<sup>[1]</sup> Diederik P. Kingma, Jimmy Ba - Adam: A Method for Stochastic Optimization

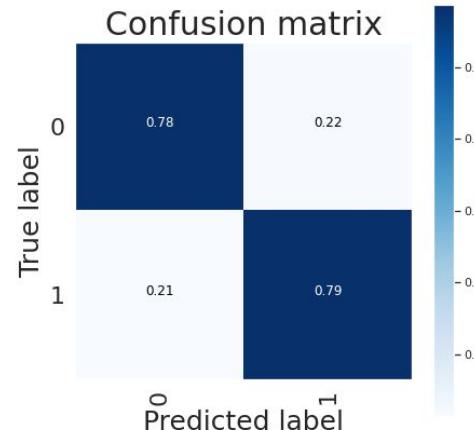




150 LSTM Units - 0.3 Dropout



200 LSTM Units - 0.5 Dropout



Label: Positive

Score: 0.8393939733505249

Time Elapsed: 0.24189233779907227 sec

Happy and relieved to have the #PfizerBioNTech #Covidvaccine MONEY – amazing work from all at @MHRAgovuk GPE and @NHSuk GPE since... <https://t.co/3iRxehHIDa>

Label: Negative

Score: 0.28334230184555054

Time Elapsed: 0.24546098709106445 sec

So, Trump PRODUCT spent \$12.4 Billion MONEY on # OperationWarpSpeed MONEY to speed # CARDINAL vaccine development and distribution. After all the... <https://t.co/zdvBE4VSEu>

Label: Negative

Score: 0.45411166548728943

Time Elapsed: 0.24692249298095703 sec

By 8 p.m. TIME Sunday DATE, more than 93,000 CARDINAL people in # HongKong PERSON had received the first ORDINAL dose of #COVID19 vaccines, 91,800 CARDINAL of... <https://t.co/gpcQy6b2UX>

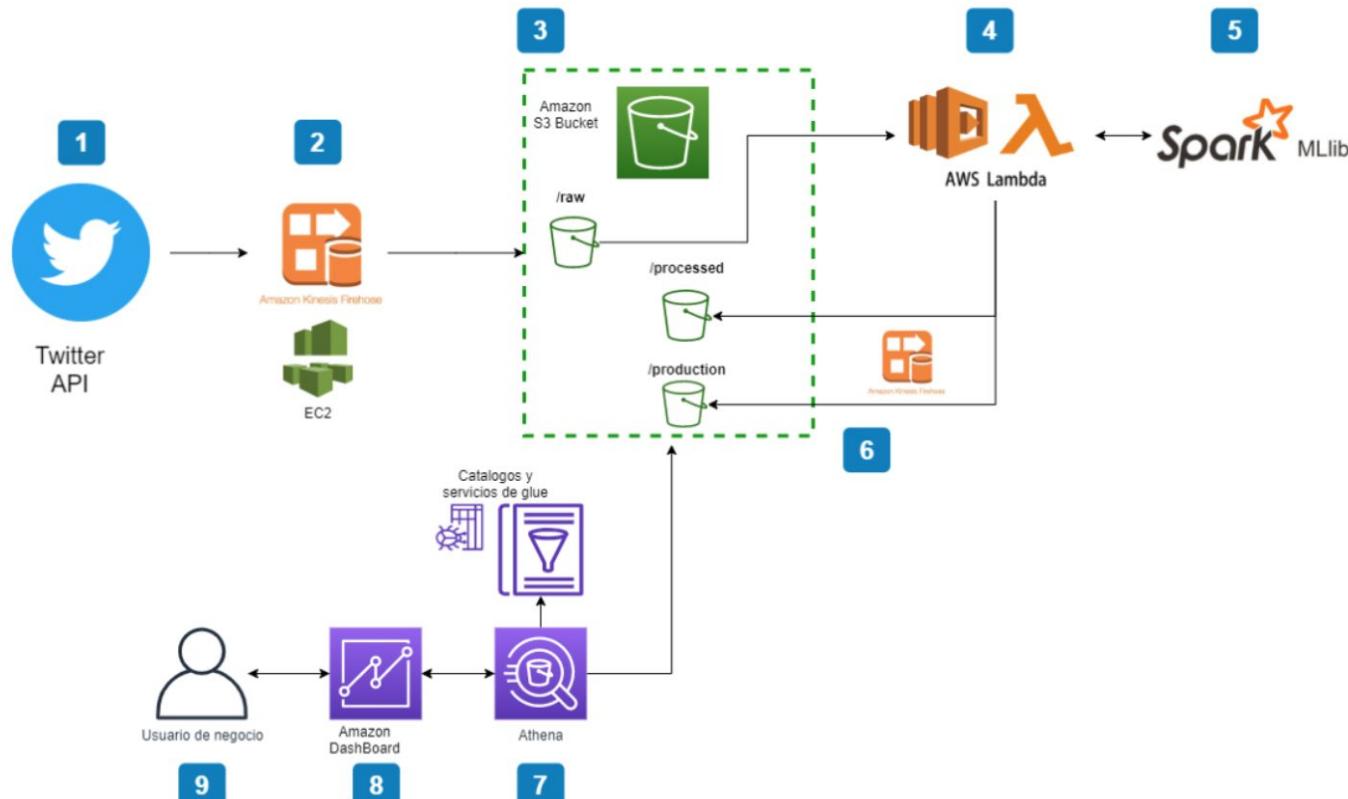
Label: Positive

Score: 0.7704590559005737

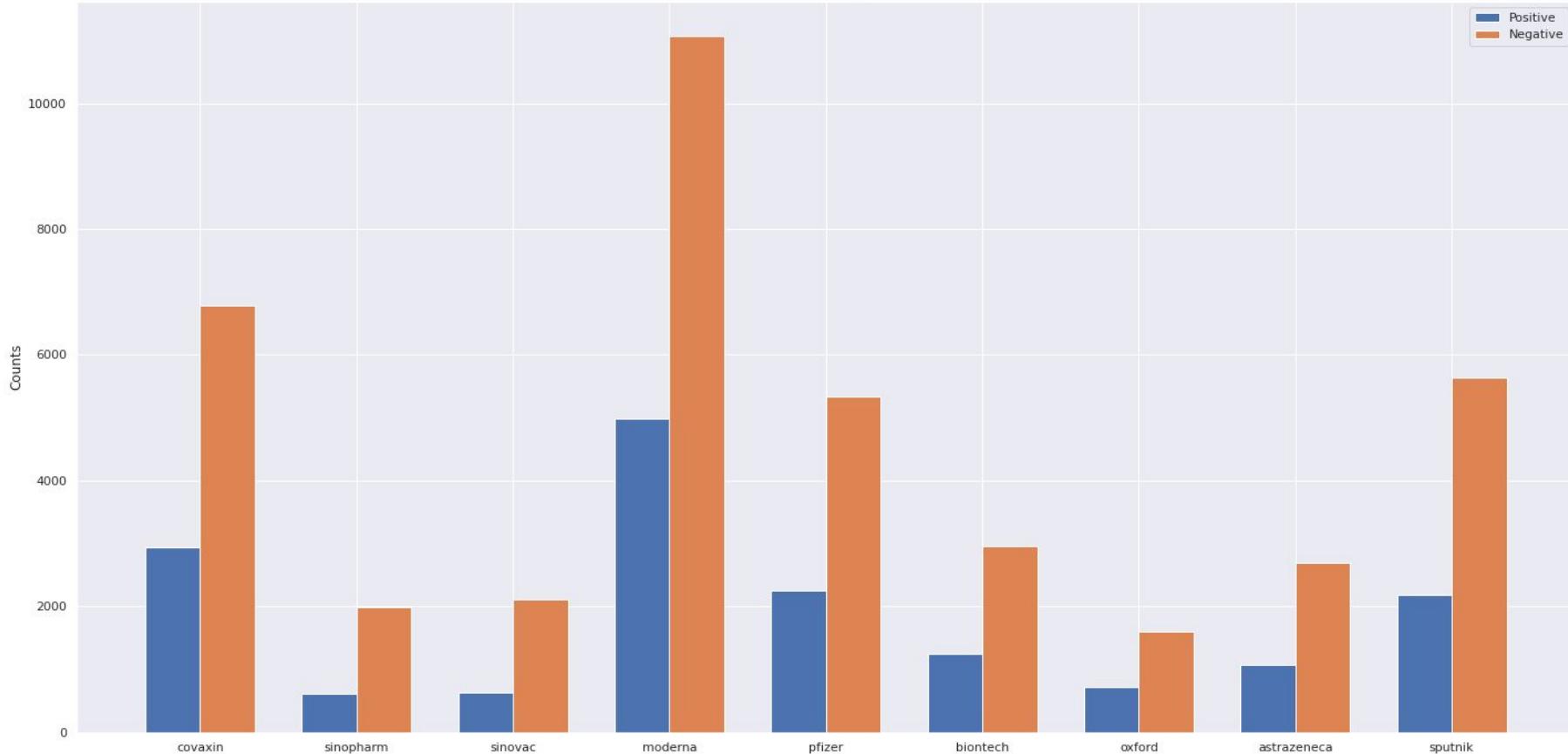
Time Elapsed: 0.2669246196746826 sec

Minutes after the POTUS tweet, Vice Pres. Pence PERSON said the FDA ORG could issue EUA to Moderna GPE within hours TIME.

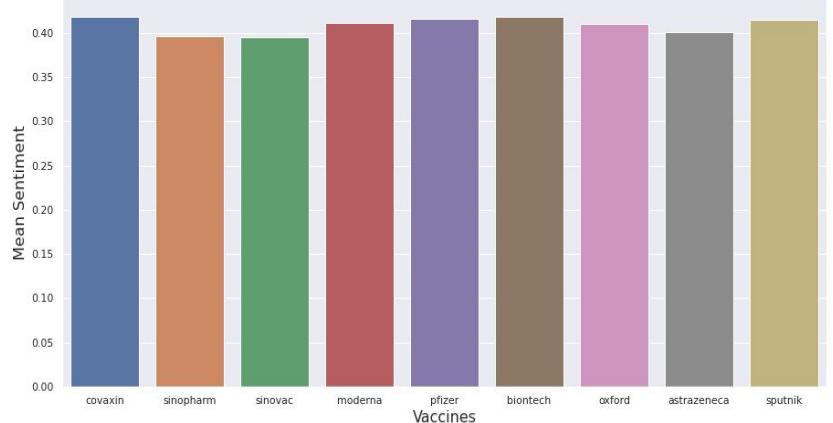
Pence PERSON was s... <https://t.co/l0y1MQo5vY>



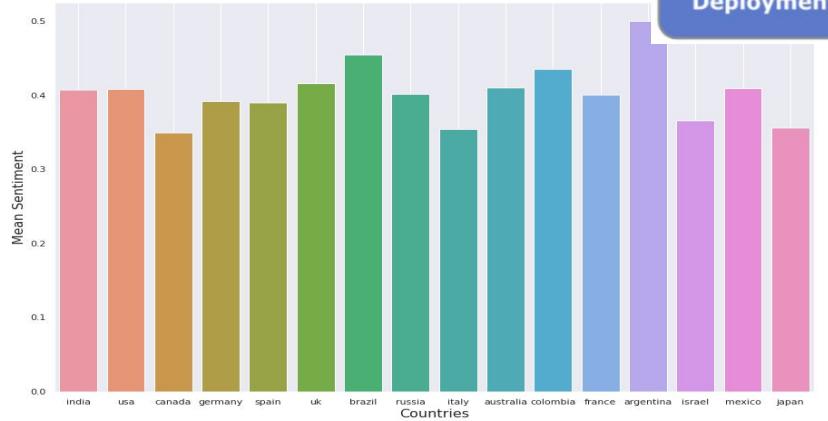
Sentiment Counts by Vaccine

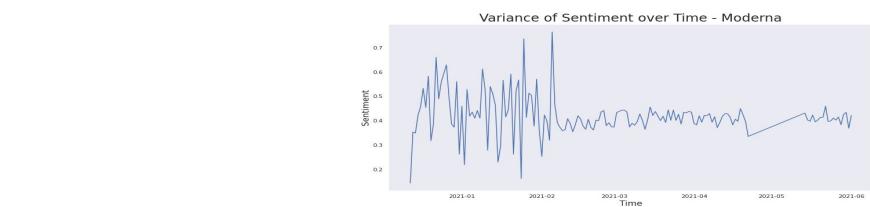
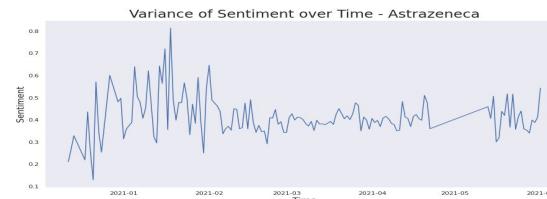
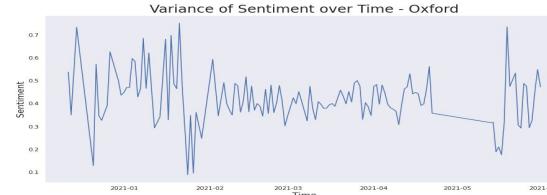
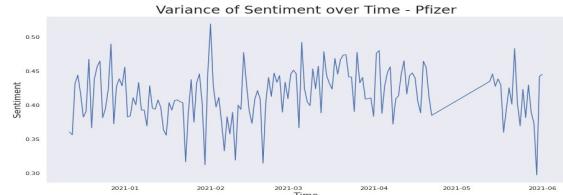


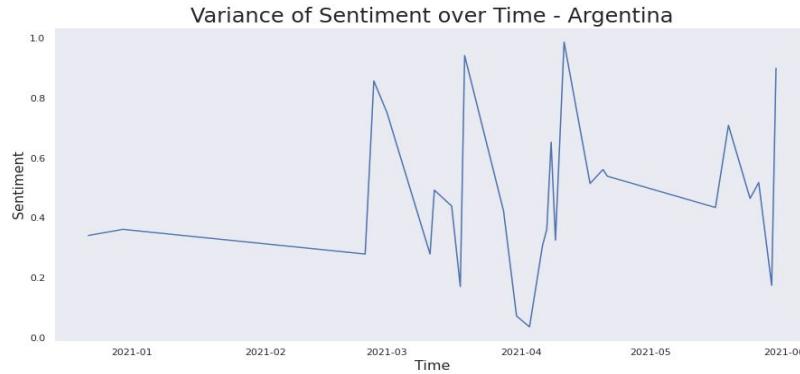
## Mean Sentiment of Vaccines



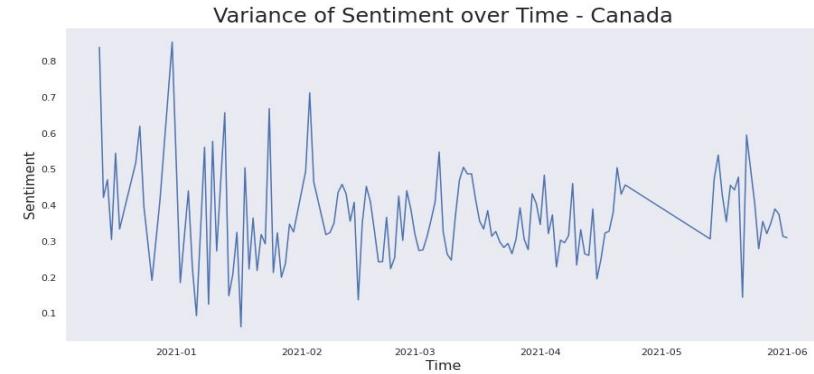
## Mean Sentiment of Countries



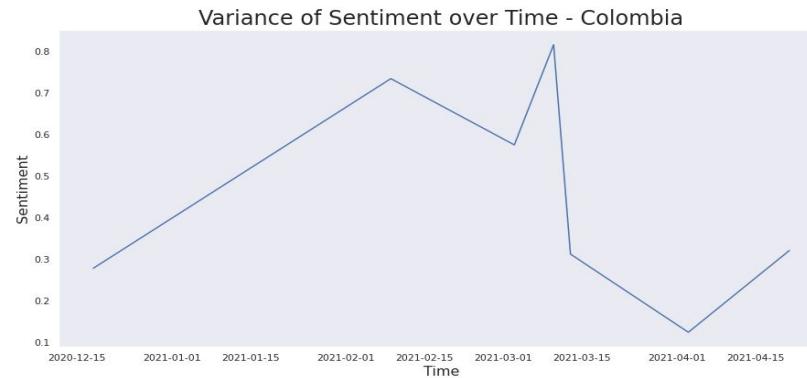




Most Positive Country



Most Negative Country



PHASE	CRISP-DM PLANNING															
	WEEK 1					WEEK 2					WEEK 3					
	M	T	W	R	F	M	T	W	R	F	M	T	W	R	F	
Business Understanding																
Data Understanding																
Data Preparation																
Modeling																
Evaluation																

 Code

 Issues

 Pull requests

 Actions

 Projects

 Wiki

 Security

 Insights

 main 

 1 branch

 0 tags

 Go to file

 Add file 

 Code 



ctezna Update README.md

37a7243 40 minutes ago  6 commits

 data

Update README.md

1 hour ago

 notebooks

Add files via upload

1 hour ago

 README.md

Update README.md

40 minutes ago

README.md

## Covid Vaccine NLP Analysis

Steps to reproduce:

1. Go to data/ and download datasets from Kaggle (rename the Twitter Sentiment 140 dataset to twitterSentiment.csv)
2. Create a bucket in Amazon S3 that will contain the datalake. For example "docstore-datalake".
3. Inside "docstore-datalake" create 5 zones as follows:  


For more information on how to reproduce this project please visit the project repository:  
<https://github.com/ctezna/covidVaccine-nlp>