Predicting Bitcoin Movements by Analyzing News

By: Ata Akca

Index

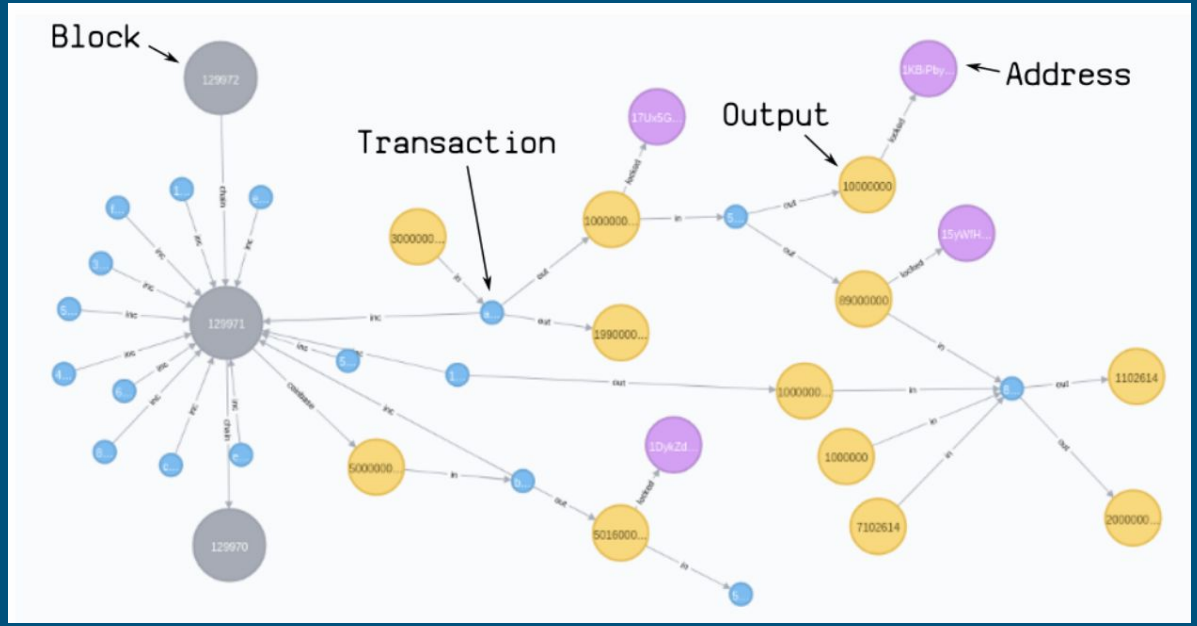
- Problem Statement
- External Research / Background Information
- Data Scraping
- Data Cleaning / Eda
- Word Vectorizing and NLP
- Modeling and Results
- What’s next
- Questions

Problem Statement

Are bitcoin movements predictable by looking at the news?

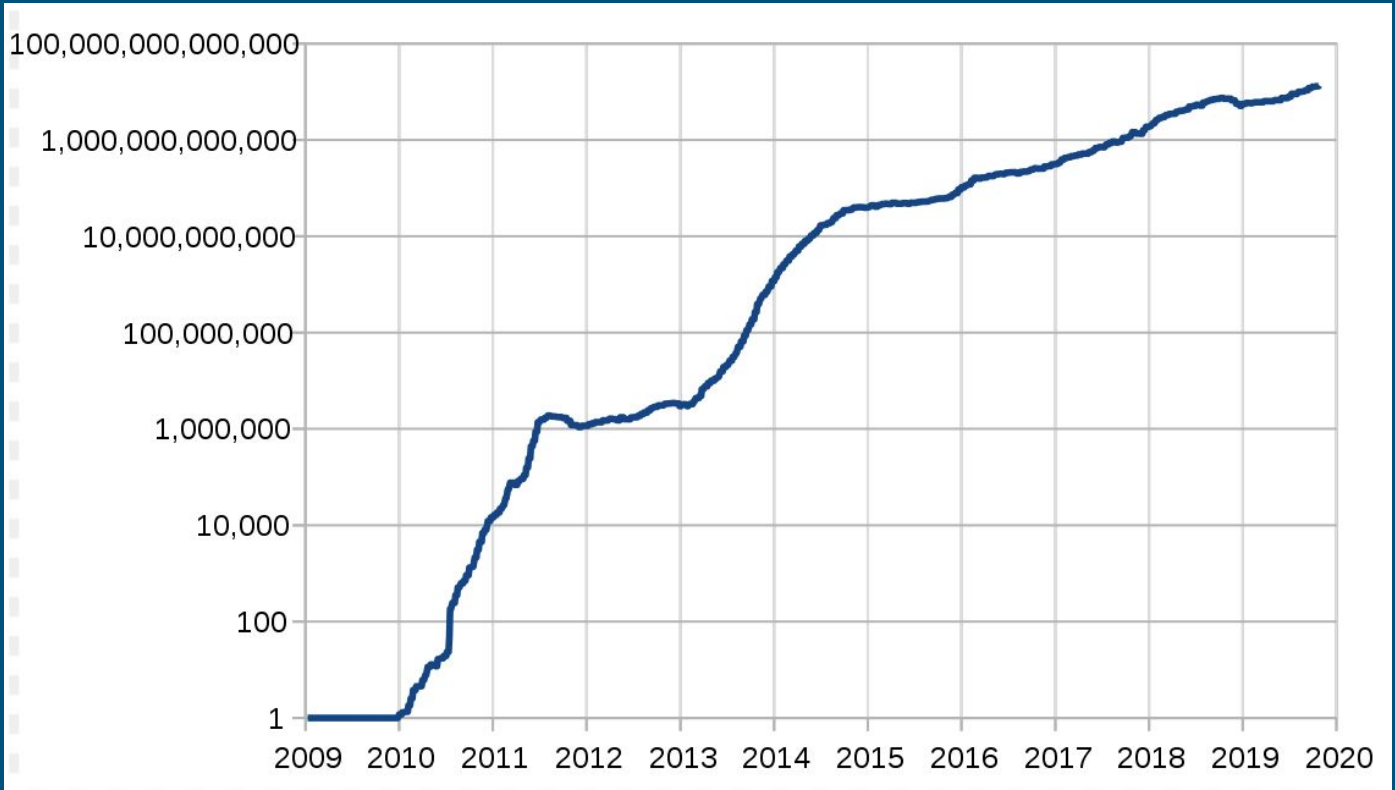
What is Bitcoin?

- Digital Currency
- Blockchain
- Digital Ledger
- Halving



What Block Chain Looks Like

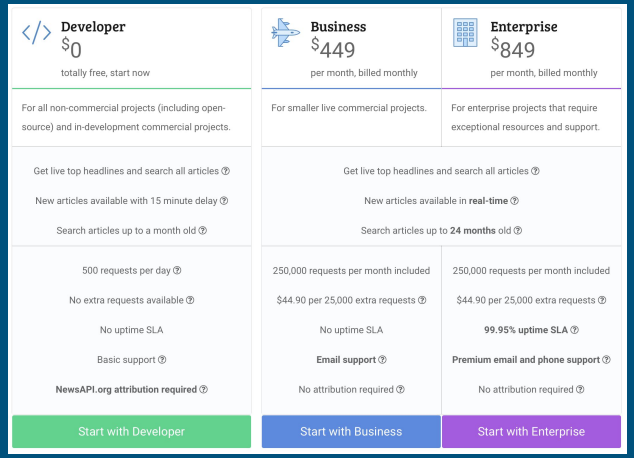
Bitcoin Mining Difficulty

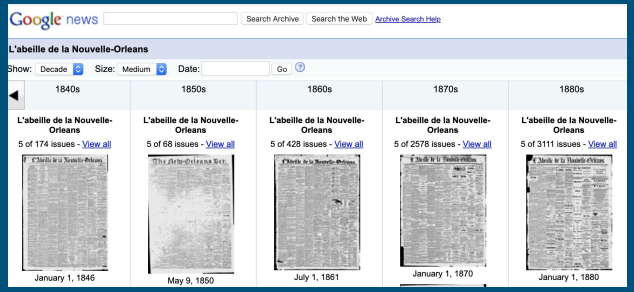


Data Scraping

- Selenium
- ChromeDriverManager
- Merging
- Saving
- Non-Viable Alternatives:
  - Wikipedia (List of Online News)
  - News API
  - Google News

The New York Times





The New York Times - Search
nytimes.com/search?dropmab=true&endDate=20200228&query=&sections=World%7Cnyt%3A%2F%2Fsection%2F70e865b6-cc70-5181-8...
Chrome is being controlled by automated test software.

Feb. 28

ASIA PACIFIC

Just One American With Coronavirus Still in Hospital: Vice President Pence

Of dozens of confirmed coronavirus cases in the United States, only one patient is still in the hospital, U.S. Vice President Mike Pence said on Friday.

By Reuters

Feb. 28

EUROPE

Italy's Quarantine Leaves Locked-Down Towns Feeling 'Sacrificed'

Friends and families have been divided and businesses shuttered. But care packages of focaccia and cigarettes ease the strain.

By Jason Horowitz and Emma Bubola

PRINT EDITION Italian Towns Feel 'Sacrificed' by Quarantine | February 29, 2020, Page A9

Feb. 28

EUROPE

Cyprus Says Turkey Deliberately Sent It Migrants for Months

Cyprus' interior minister on Friday accused Turkey of deliberately channeling migrants for months to the ethnically-divided island's Greek Cypriot south to change the composition of its population.

By The Associated Press

Feb. 28

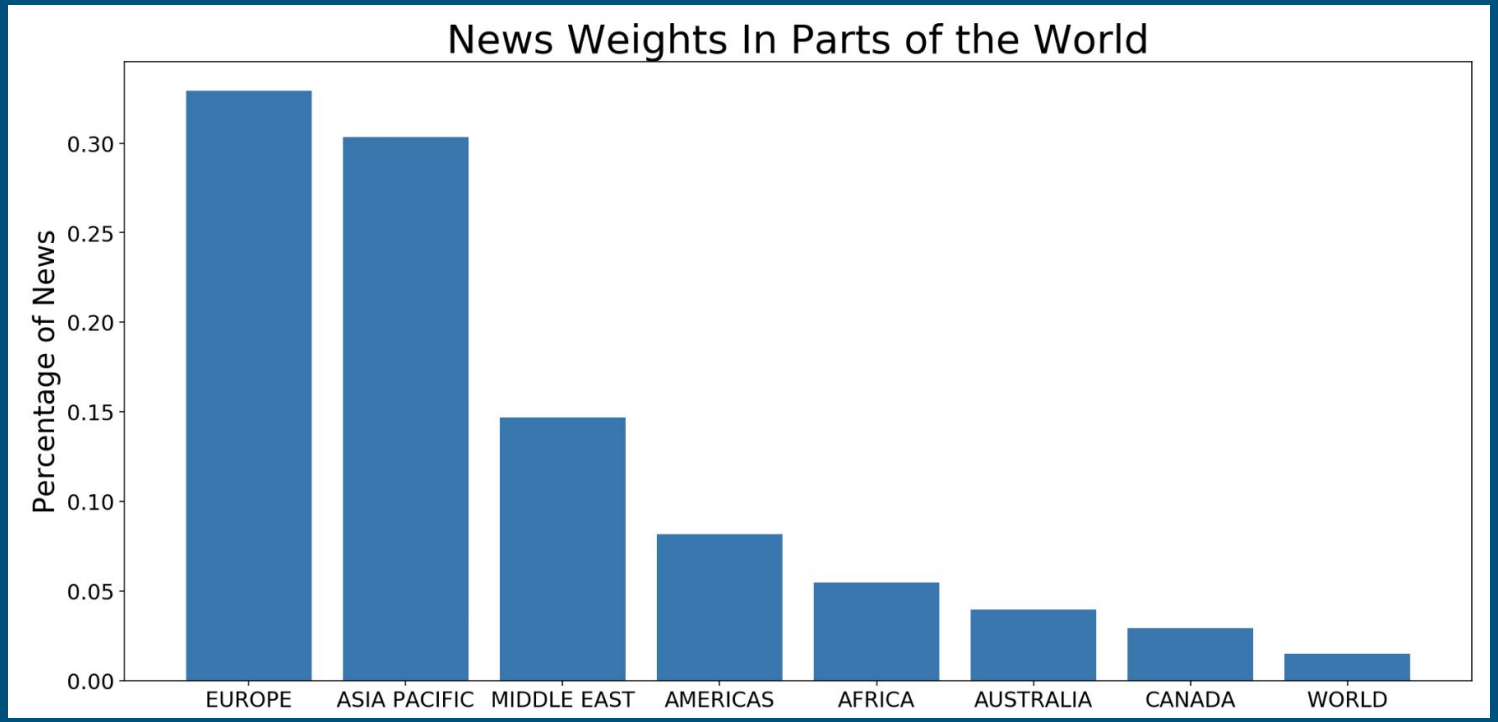
MIDDLE EAST

Confusion and Contradiction Reign in Iran's Coronavirus Reporting

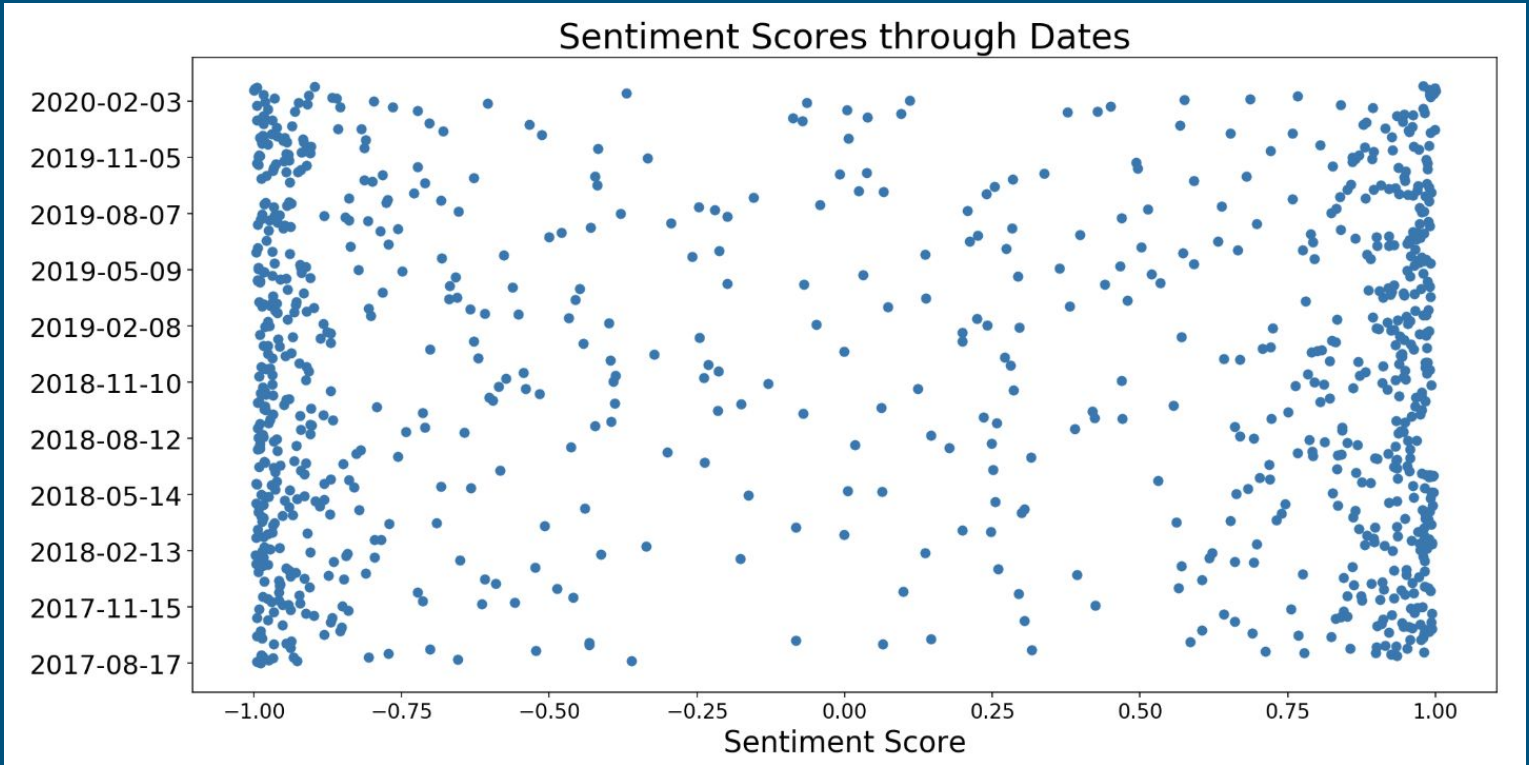
Iranian officials reported jumps in cases and deaths but denied a BBC Persia report that fatalities totaled at least 210, far higher than the

ytimes.com/2020/02/28/world/.../coronavirus-iran-confusion.html?se...

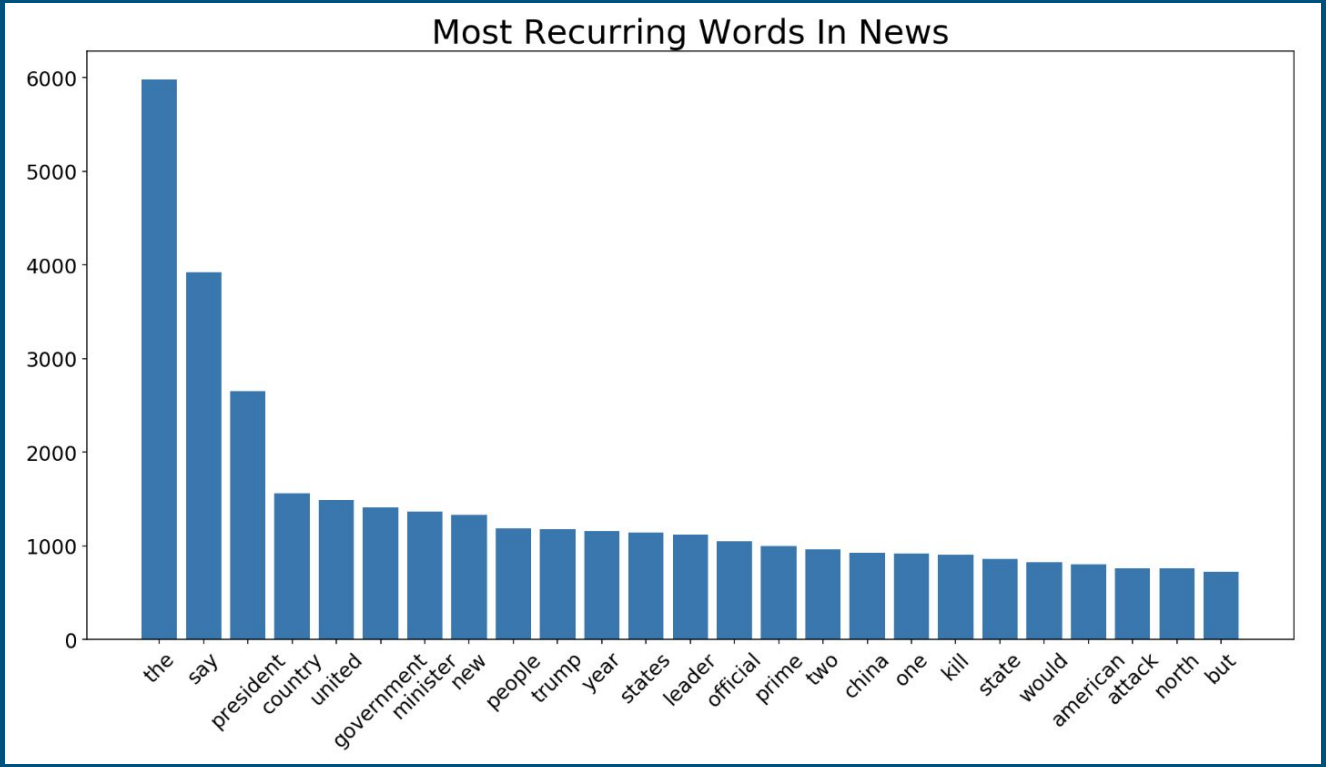
Data Cleaning / EDA



Data Cleaning / EDA

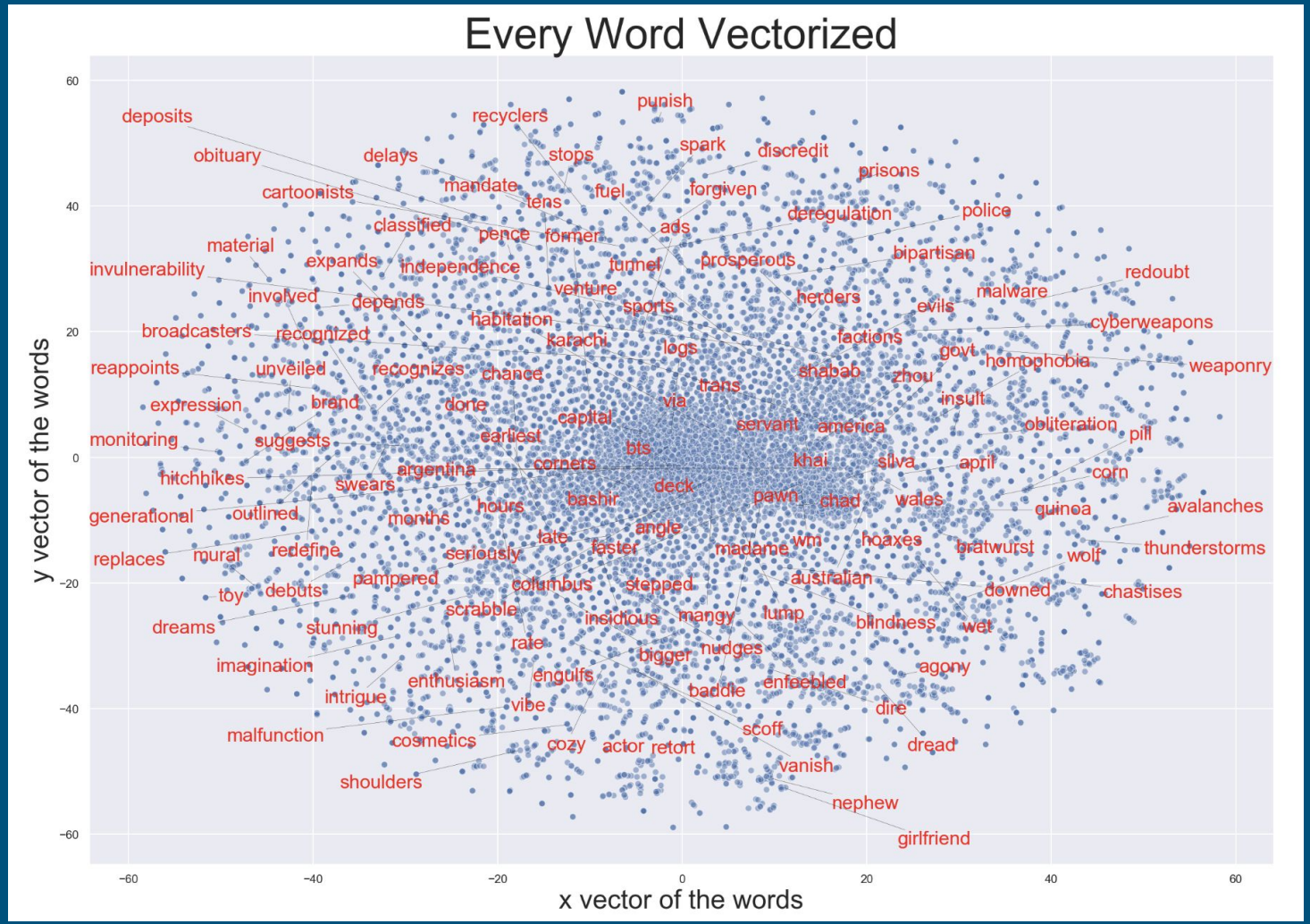


Further EDA

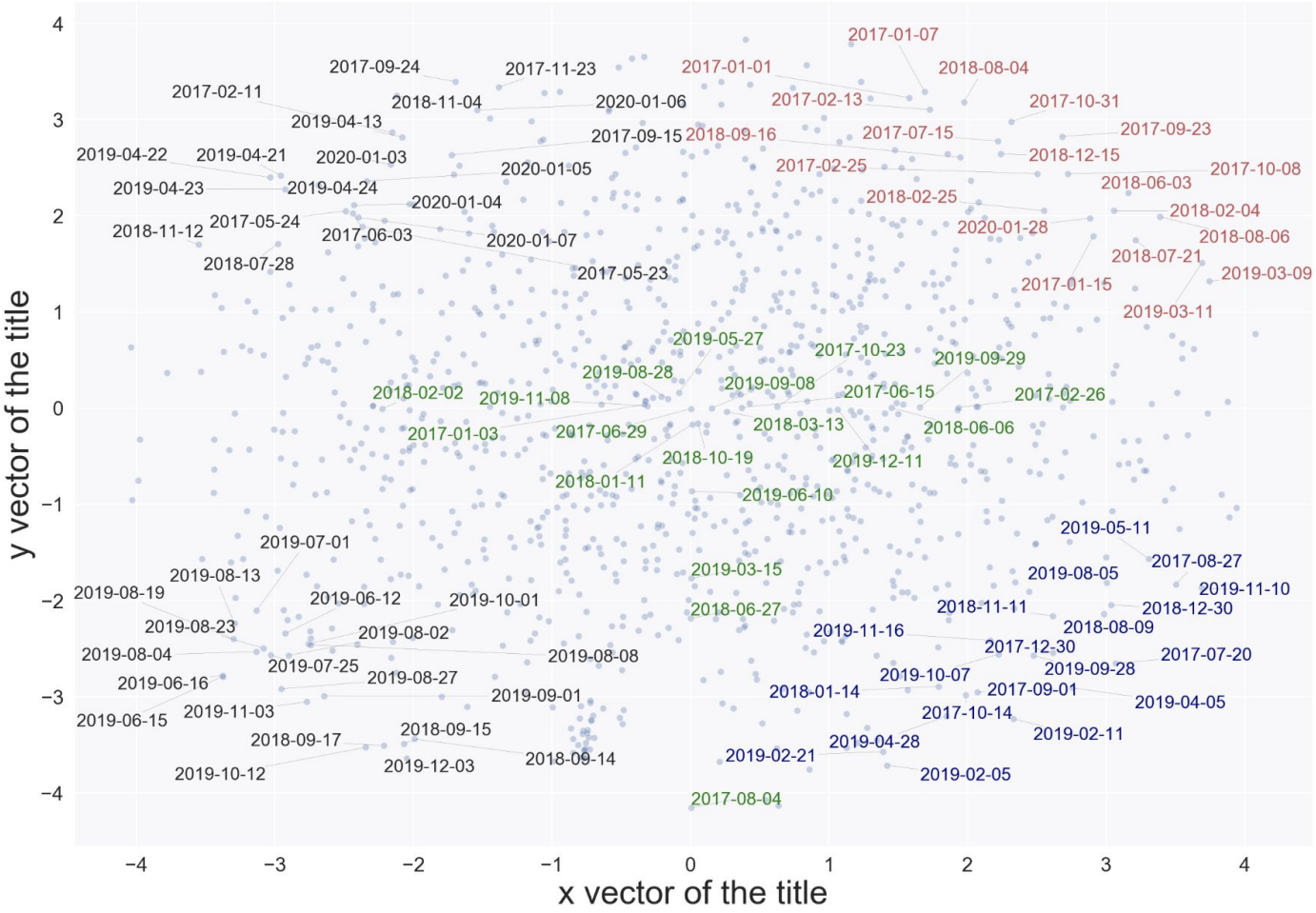

Most Recurring Words In News
6000
5000
4000
3000
2000
1000
0
the
say
president
country
united
government
minister
new
people
trump
year
states
leader
official
prime
two
china
one
kill
state
would
american
attack
north
but

Word Vectorizing

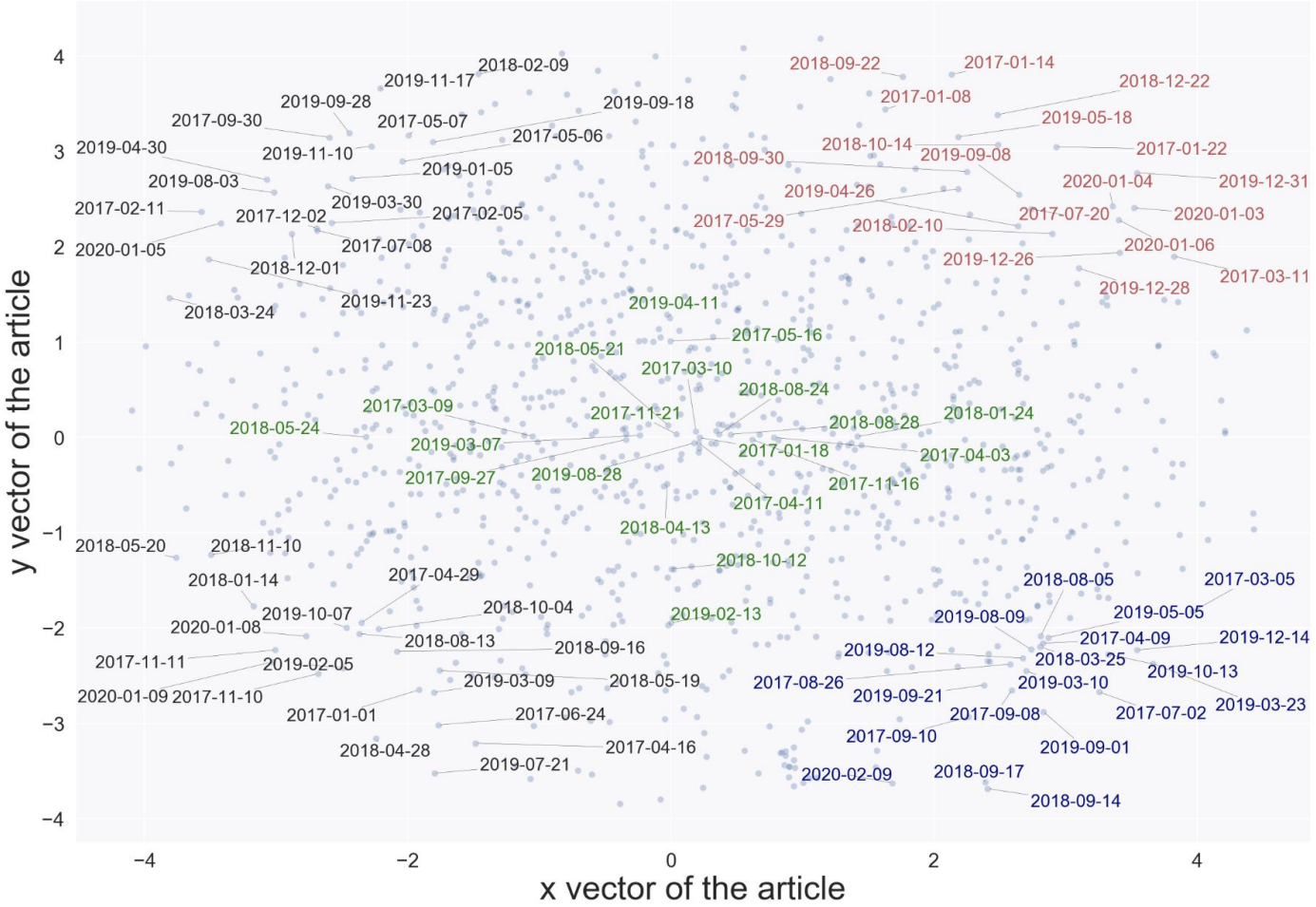
- Tokenization
- Stopwords (English)
- Google Lexicon (100B +)
    - 3.7% of words not available


Every Word Vectorized
y vector of the words
x vector of the words
deposits
recyclers
punish
obituary
delays
stops
spark
discredit
prisons
cartoonists
mandate
tens
fuel
forgiven
deregulation
police
classified
pence
former
ads
material
bipartisan
invulnerability
expands
independence
turmoil
prosperous
redoubt
involved
depends
venture
sports
herders
malware
evils
broadcasters
recognized
habitation
karachi
logs
factions
cyberweapons
reappoints
unveiled
recognizes
chance
shabab
zhou
govt
homophobia
weaponry
expression
brand
dorte
trans
via
insult
capital
servant
america
obliteration
monitoring
suggests
earliest
bts
pill
hitchhikes
argentina
corners
khai
silva
april
corn
swears
deck
pawn
chad
wales
generational
outlined
months
hours
bashir
quinoa
avalanches
replaces
mural
redefine
seriously
late
faster
angle
madame
wm
hoaxes
bratwurst
wolf
thunderstorms
pampered
columbus
stepped
australian
toy
debuts
downed
chastises
dreams
stunning
scrabble
insidious
mangy
lump
blindness
wet
rate
bigger
nudges
imagination
agony
enthusiasm
engulfs
enfeebled
intrigue
baddie
vibe
dire
malfunction
cosmetics
cozy
actor
retort
scoff
dread
vanish
shoulders
nephew
girlfriend
60
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60

Vectorized Titles

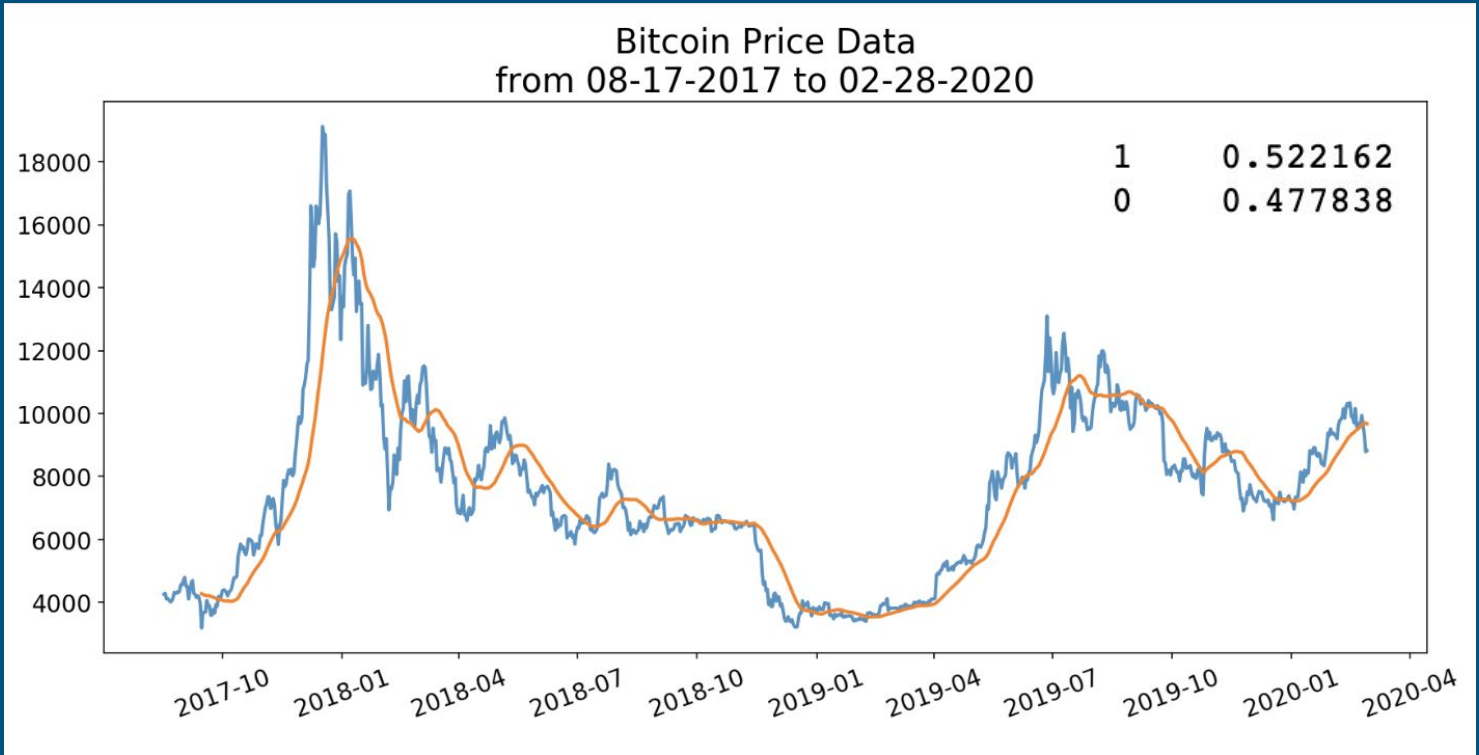


Vectorized Articles

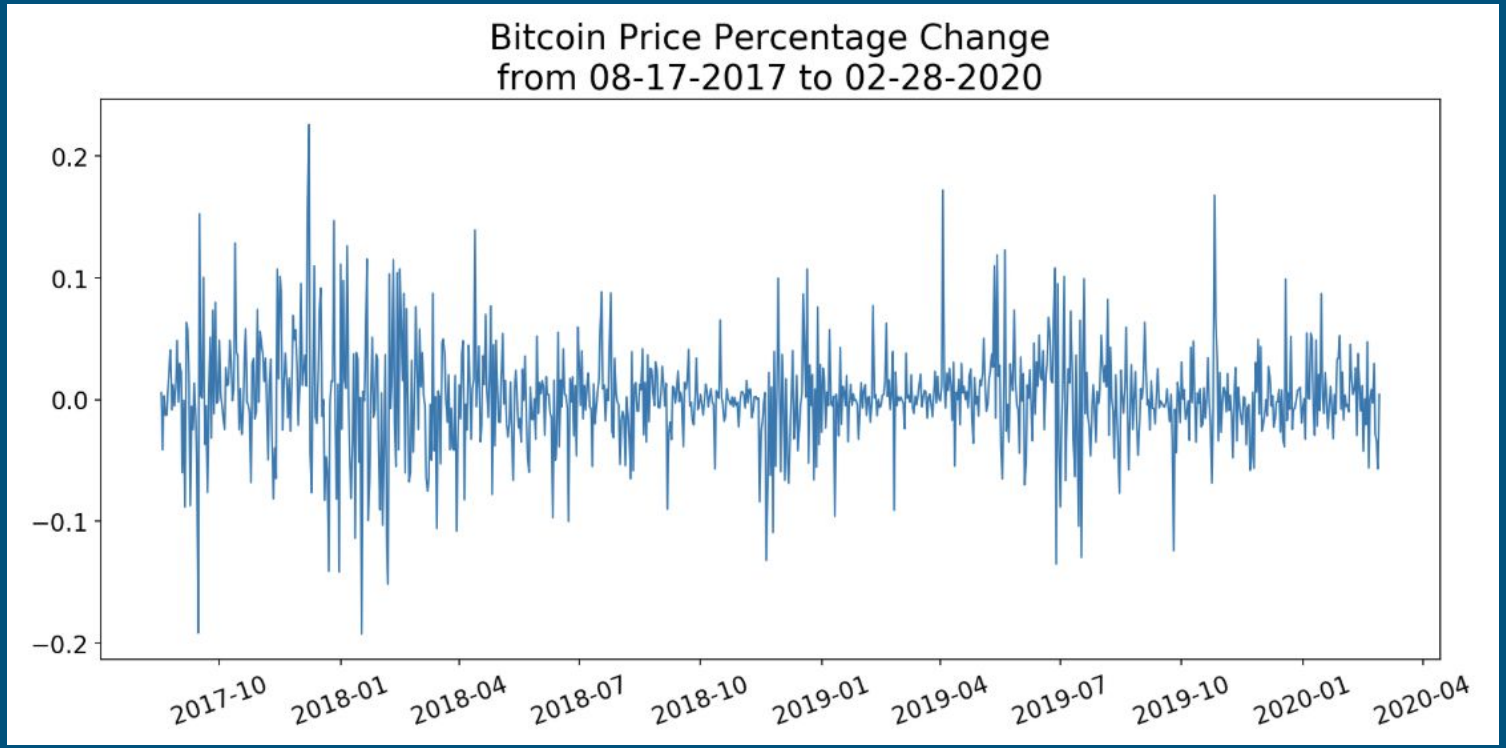


Bitcoin Price Data

ALPHA VANTAGE



Bitcoin Price Change

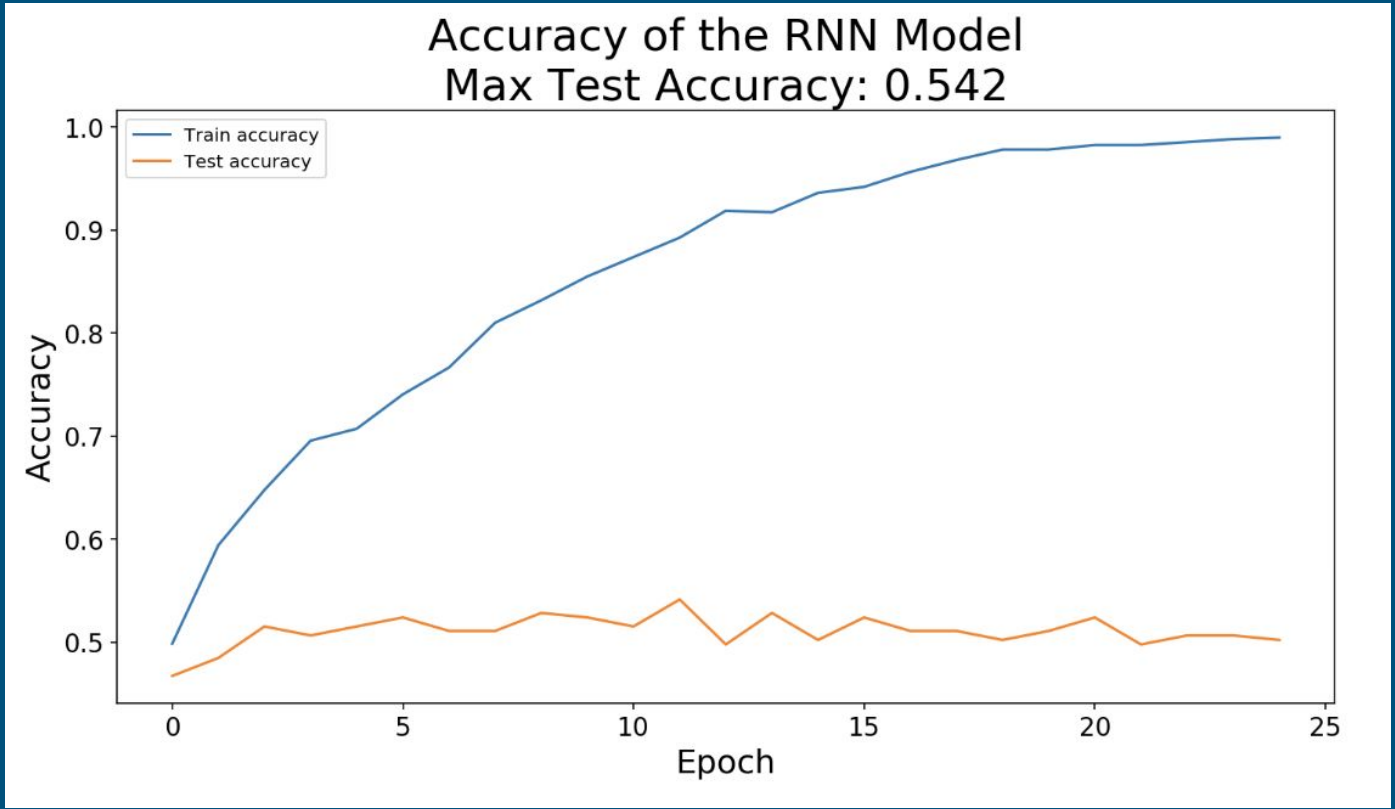


Vectorized Samples

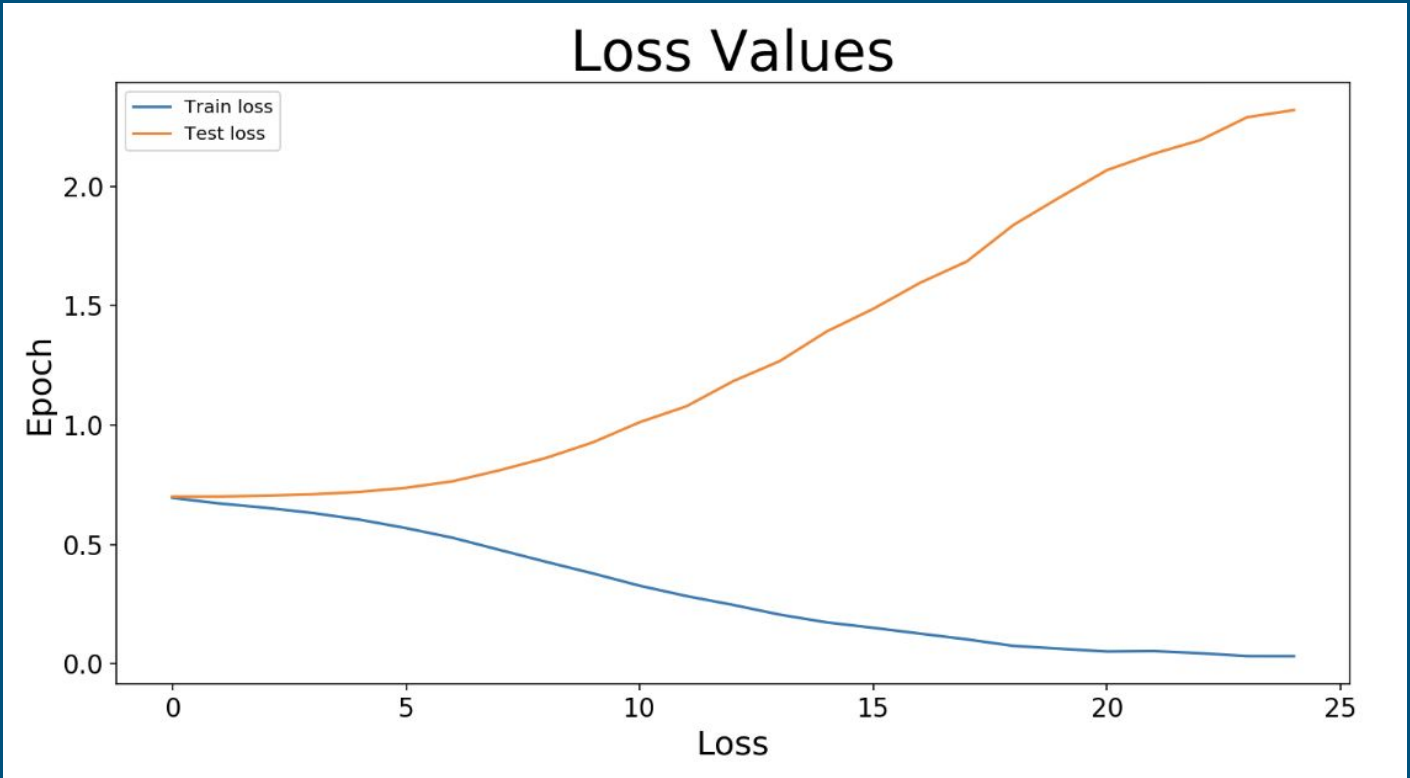
date	0	1	2	3	4	...	296	297	298	299	change
2017-08-18	0.020815	0.054416	0.046230	0.064621	-0.027624	...	-0.020247	-0.040255	0.056129	0.030431	1
2017-08-19	0.023122	0.063057	0.053862	0.026144	0.017412	...	-0.031177	-0.051869	0.058205	-0.016258	0
2017-08-20	0.028526	0.050478	0.043822	0.057757	-0.048699	...	-0.005911	-0.048312	0.043182	0.016094	1
2017-08-21	0.023134	0.059504	0.044987	0.067485	-0.044458	...	-0.029511	-0.037497	0.066254	0.006955	0
...	...	...	...	...	...	...	...	...	...	...	...
2020-02-25	0.010662	0.050200	0.033292	0.050159	-0.029390	...	-0.012283	-0.055650	0.043644	0.027952	0
2020-02-26	0.017747	0.059480	0.024659	0.061293	-0.040458	...	-0.009383	-0.043335	0.033137	0.024569	0
2020-02-27	0.023614	0.044132	0.018687	0.041366	-0.044358	...	-0.004942	-0.039867	0.047979	0.010003	0
2020-02-28	-0.008888	0.050875	0.043920	0.117509	-0.040603	...	-0.068974	-0.037455	0.065725	0.017077	1

925 rows × 301 columns

Results


Accuracy of the RNN Model
Max Test Accuracy: 0.542
Train accuracy
Test accuracy
Accuracy
Epoch
1.0
0.9
0.8
0.7
0.6
0.5
0
5
10
15
20
25

Results


Loss Values
Train loss
Test loss
Epoch
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
Loss

What To Do? How to Improve?

- More Finance Related News?
- Category Density and Clustering
- Handling Vectors
  - Missing Bitcoin Vector
  - Summing Vectors
- Detecting Sudden Shocks and Worldwide News
- Adding Features
- Testing the model with new untrained News

Questions?

Thank you