Fake News detection

Author: Rüdiger Hass

TABLE OF CONTENTS

01

Introduction

Understanding the problem


PROPAGANDA

Fake News

"Fake news, [...] is a form of news consisting of deliberate disinformation or hoaxes spread via traditional news media (print and broadcast) or online social media." (Wikipedia).

Problems with Fake News

misinformation
&
manipulation

"complex, both technically and philosophically."

Mark Zuckerberg (2016)

Goal

A machine learning model which can be applied to distinguish Fake News contents from true news contents by Natural Language Processing (NLP).

02
DATA

Visualizing data

SHUTTERSTOCK, CNNMONEY

DATASET

- corpus comprises output of 9 publishers in a week close to the US elections 2016

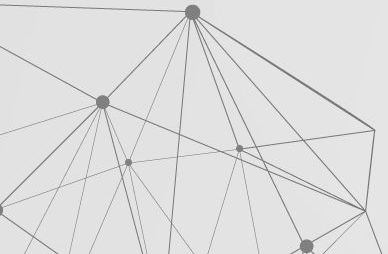
- Articles from 09/19 to 09/23 and 09/26 and 09/27/2016
- fact-checked by professional journalists at BuzzFeed

DATA SET

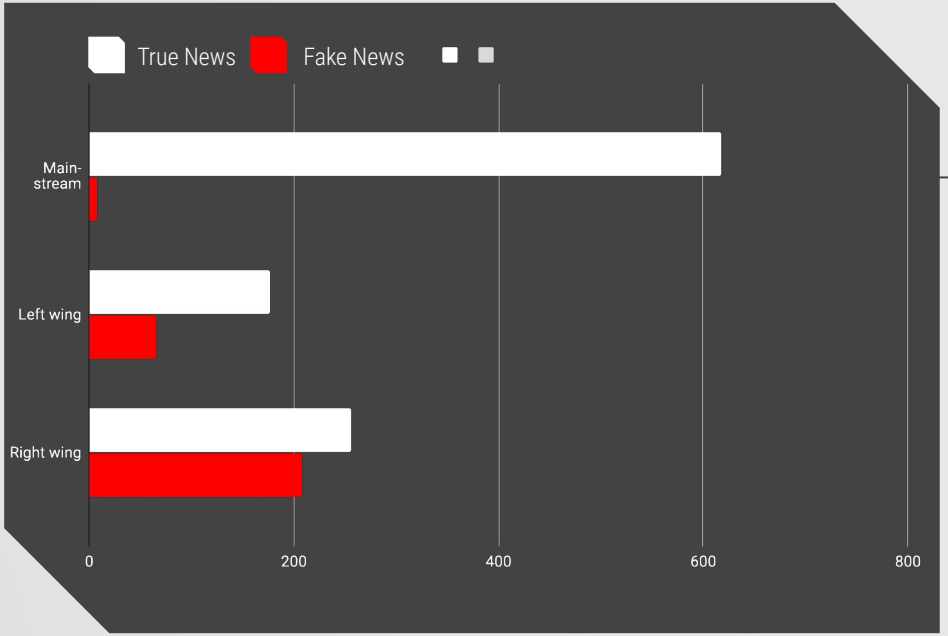
1496 Total Articles

1210 True Content

286 Fake News



Political Direction



Output of 9 publishers in a week close to the US elections. Among the selected publishers are 6 prolific hyperpartisan ones
All publishers earned Facebook's blue checkmark, indicating authenticity and an elevated status within the network

Mainstream

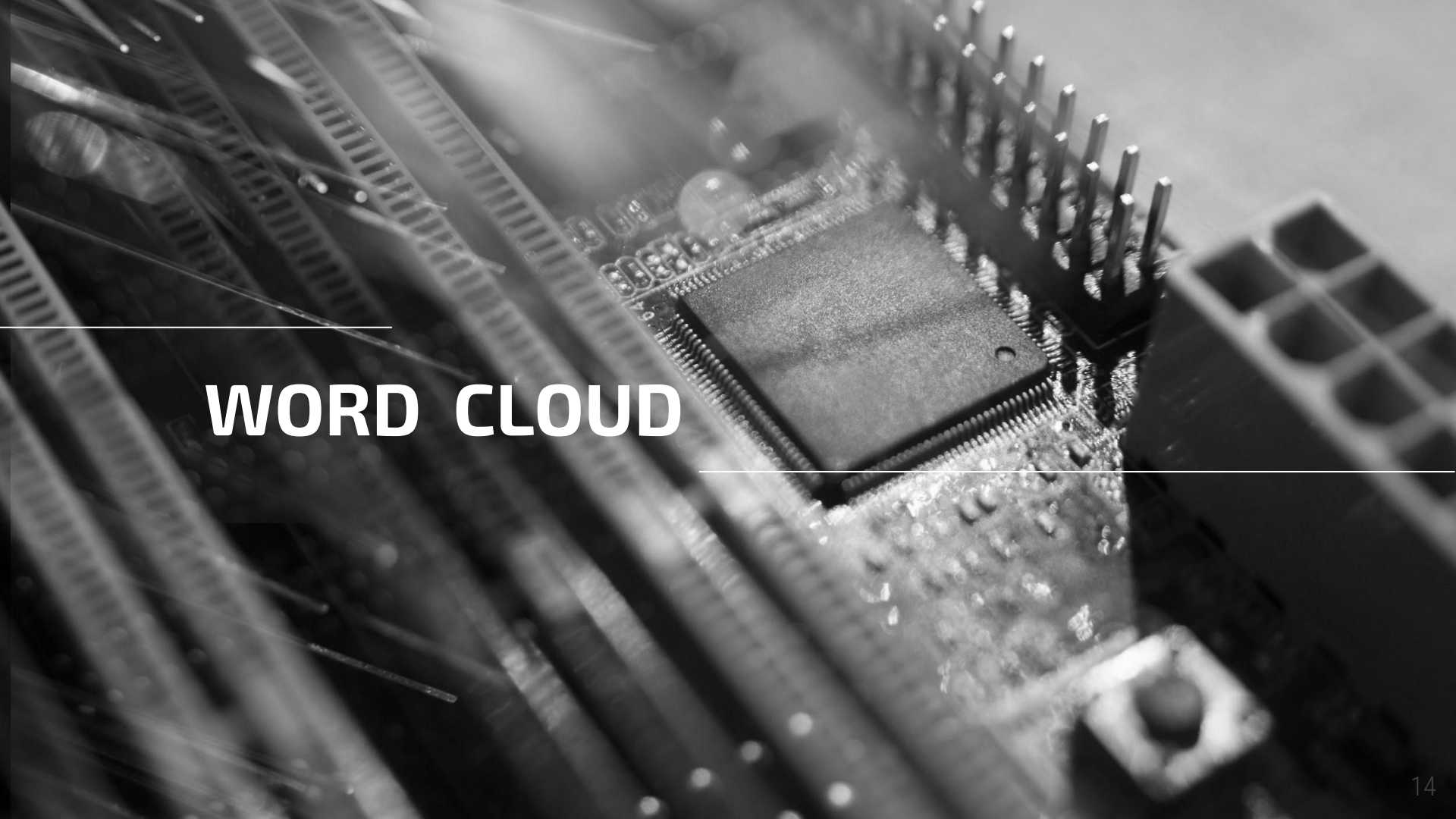
abc CNN politico

Left wing

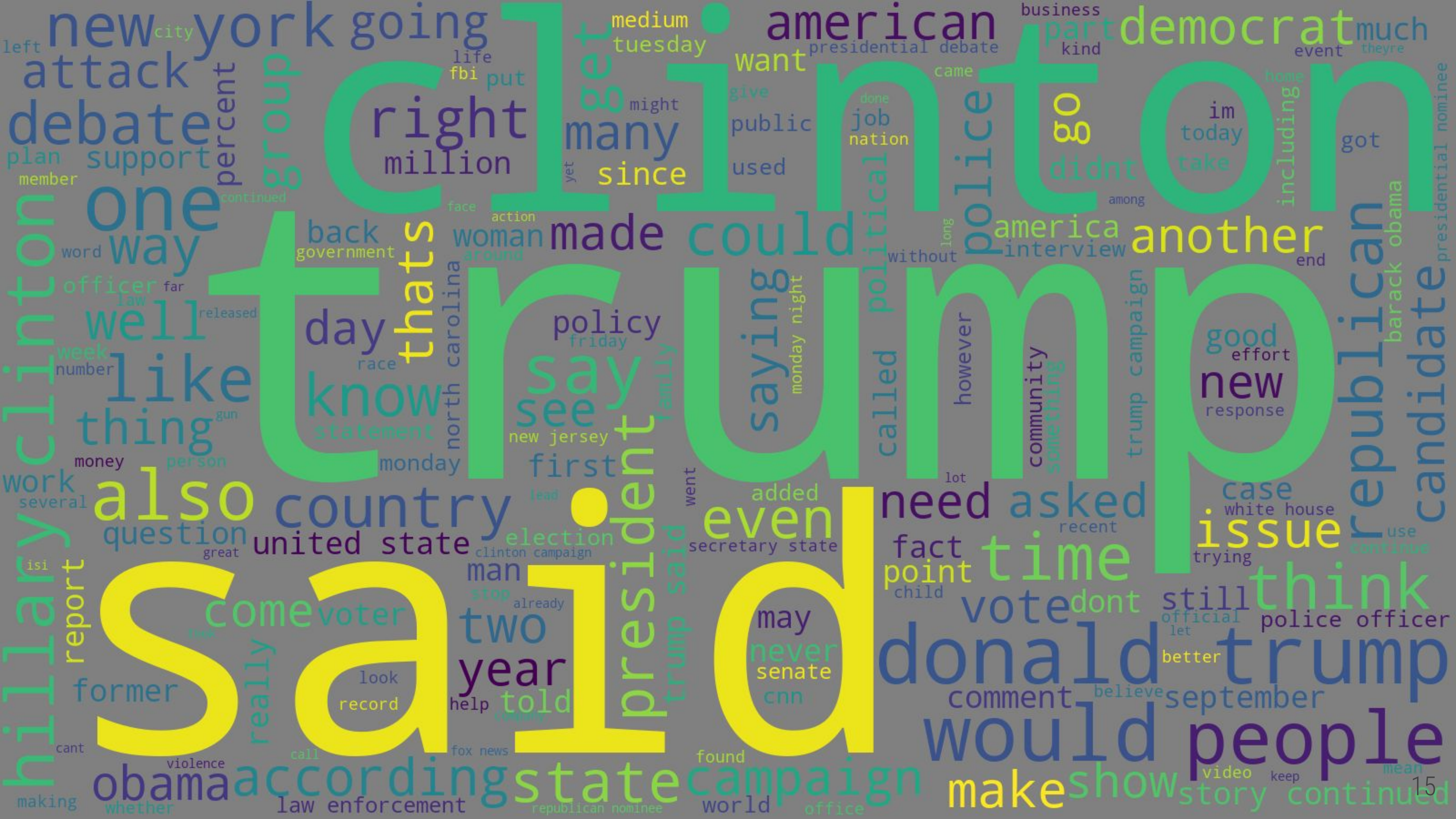
Occupy Democrats other98 addicting-info

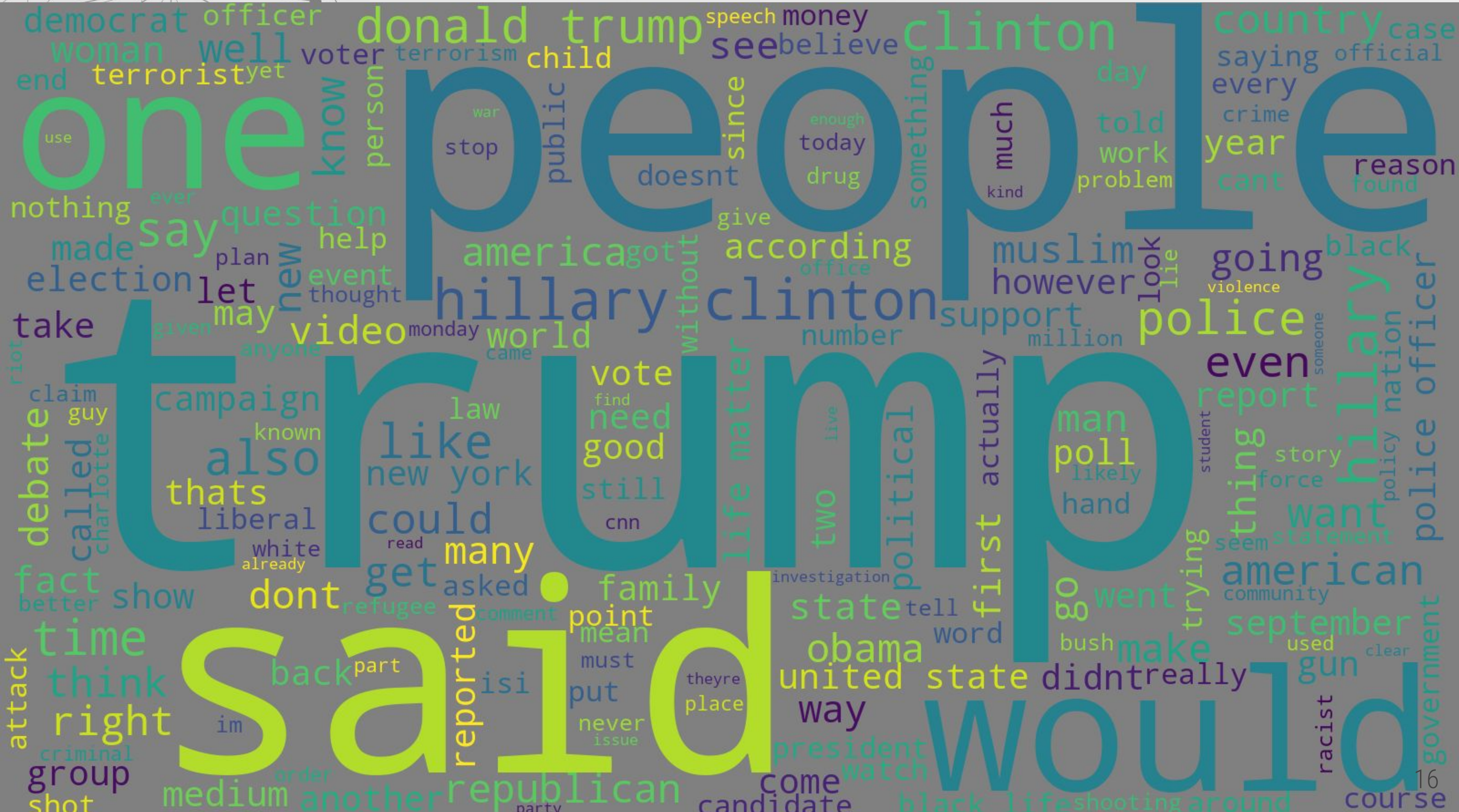
Right wing

Freedom Daily Eagle Rising Right Wing News



WORD CLOUD


clinton
trump
said
new york
going
american
democrat
much
business
medium
tuesday
presidential debate
attack
debate
right
many
public
job
nation
im
today
got
one
support
million
since
used
police
go
didnt
take
including
way
back
woman
made
could
political
america
another
interview
republican
candidate
barack obama
presidential nominee
thats
hillary clinton
well
policy
day
north carolina
saying
monday night
called
however
community
something
trump campaign
good
effort
new
response
like
know
say
see
thing
statement
new jersey
monday
first
president
need
asked
case
white house
issue
also
country
even
time
fact
point
question
united state
election
secretary state
clinton campaign
think
dont
still
official
police officer
vote
man
come
voter
two
may
never
senate
donald
trump
year
told
former
report
really
look
record
help
comment
september
believe
would
people
obama
according
state
campaign
show
make
story continued
law enforcement
world
found
video
keep
mean
fox news
violence
making
whether
republican nominee
office


people
trump
said
would
one
donald trump
clinton
hillary clinton
hillary
democrat
officer
speech
money
country
case
woman
well
voter
terrorism
child
see
believe
saying
official
end
terrorist
yet
day
every
know
person
war
enough
something
much
told
crime
use
stop
public
since
today
work
year
doesnt
drug
problem
reason
cant
found
nothing
ever
kind
question
say
give
help
made
plan
america
got
according
muslim
going
black
election
let
new
event
office
look
lie
however
violence
thought
without
take
may
given
video
monday
world
support
police
police officer
riot
anyone
came
number
million
even
someone
nation
claim
vote
find
report
campaign
law
need
man
guy
known
like
life matter
actually
student
debate
called
charlotte
also
good
poll
thing
story
new york
likely
policy
thats
still
force
political
hand
liberal
could
cnn
two
want
white
read
many
seem
statement
already
first
trying
fact
get
investigation
american
show
asked
family
community
better
dont
refugee
comment
point
state
tell
go
went
september
time
mean
word
obama
bush
make
used
clear
government
must
think
back
part
united state
didnt
really
gun
isi
theyre
put
reported
attack
right
place
way
racist
im
never
issue
criminal
president
group
order
watch
come
medium
another
republican
candidate
black life
shooting
around
course
shot
party

03
MODEL

Explaining underlying models

NATURAL LANGUAGE PROCESSING

TEXT CLEANING

Lemmatization, set lowercase,
remove: punctuations, stopwords, URL's, twitter names

TOKENIZATION

Breaking up sentences into sequences of words

TRAIN MODEL

apply data in model

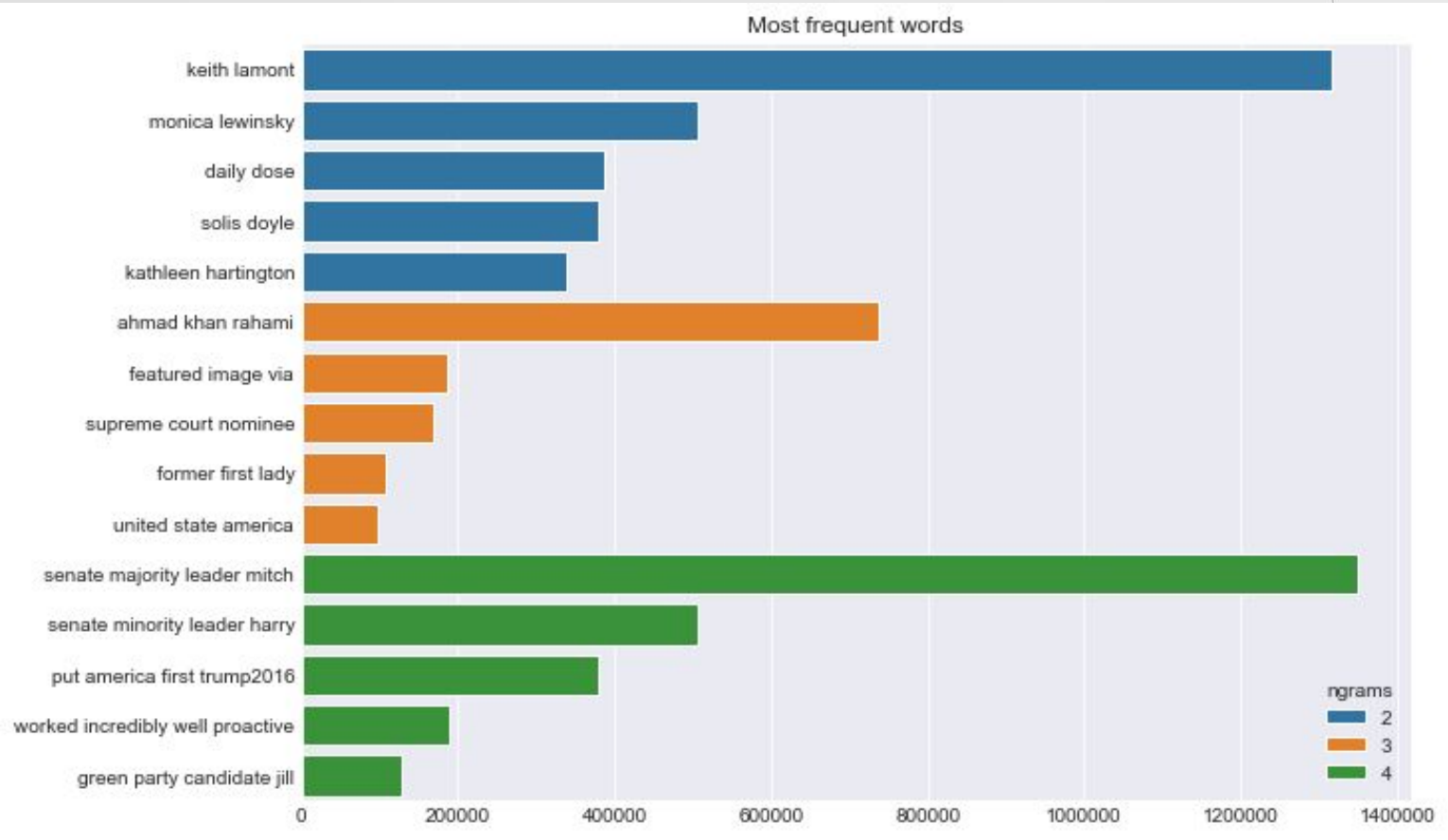
Models


MODELS
Machine Learning
Recurrent neural network
Bag-of-Words
term frequency - inverse document frequency
Word Embedding
Word2Vec

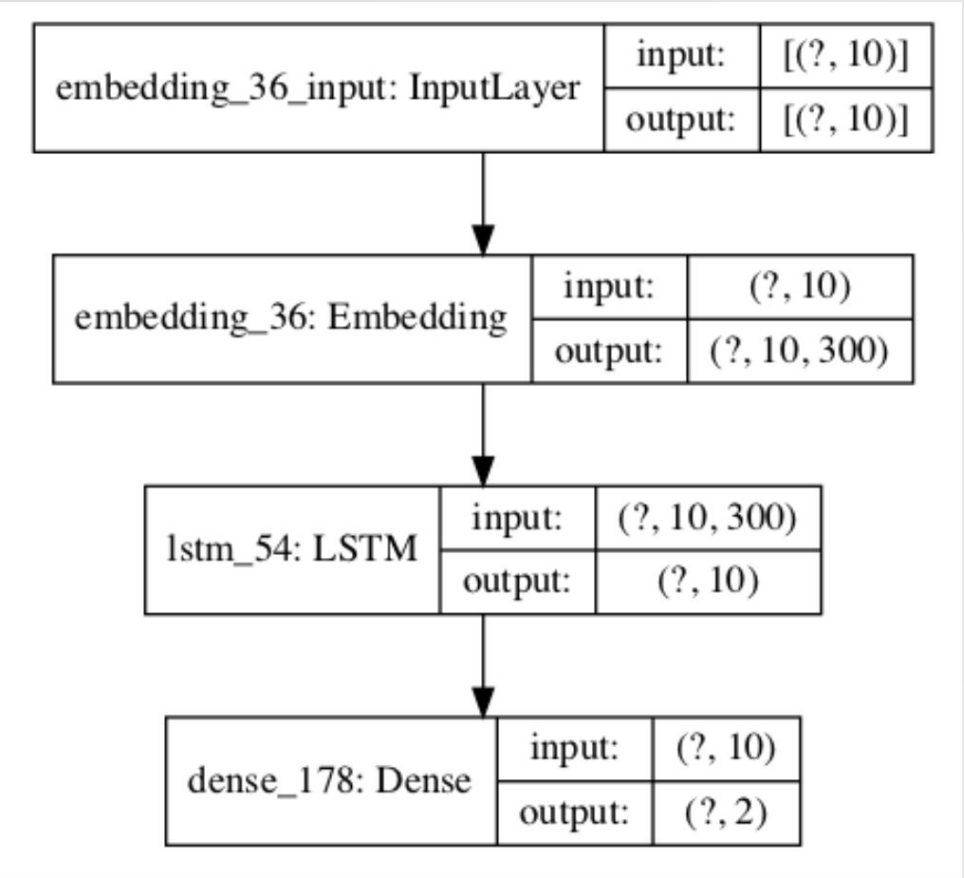
Bag-of-Words: Feature Importance

	feature	score	y
1828	combetta	0.965358	1
4088	hillary	0.963603	1
2809	drug	0.960520	1
7127	reddit	0.942785	1
7950	sharia	0.926557	1
2099	coughing	0.924741	1
8983	townsend	0.903395	1
3924	haiti	0.902220	1
1070	black	0.897204	1
2098	cough	0.888368	1

Word2Vec: ngrams



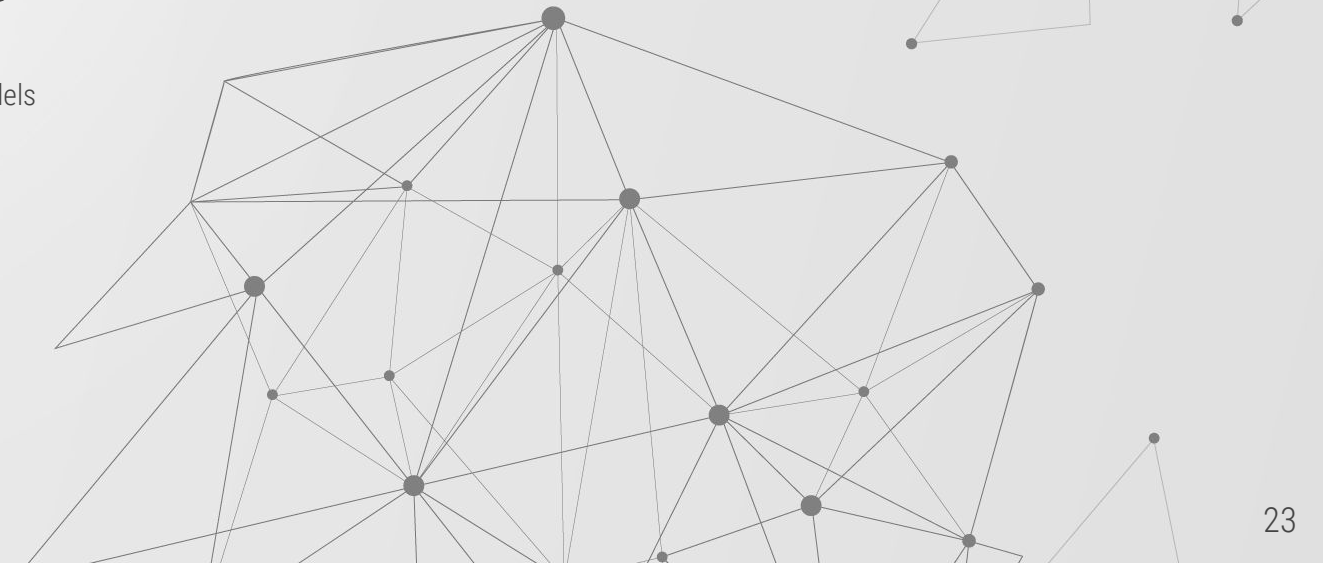
Word2Vec: Neural Network



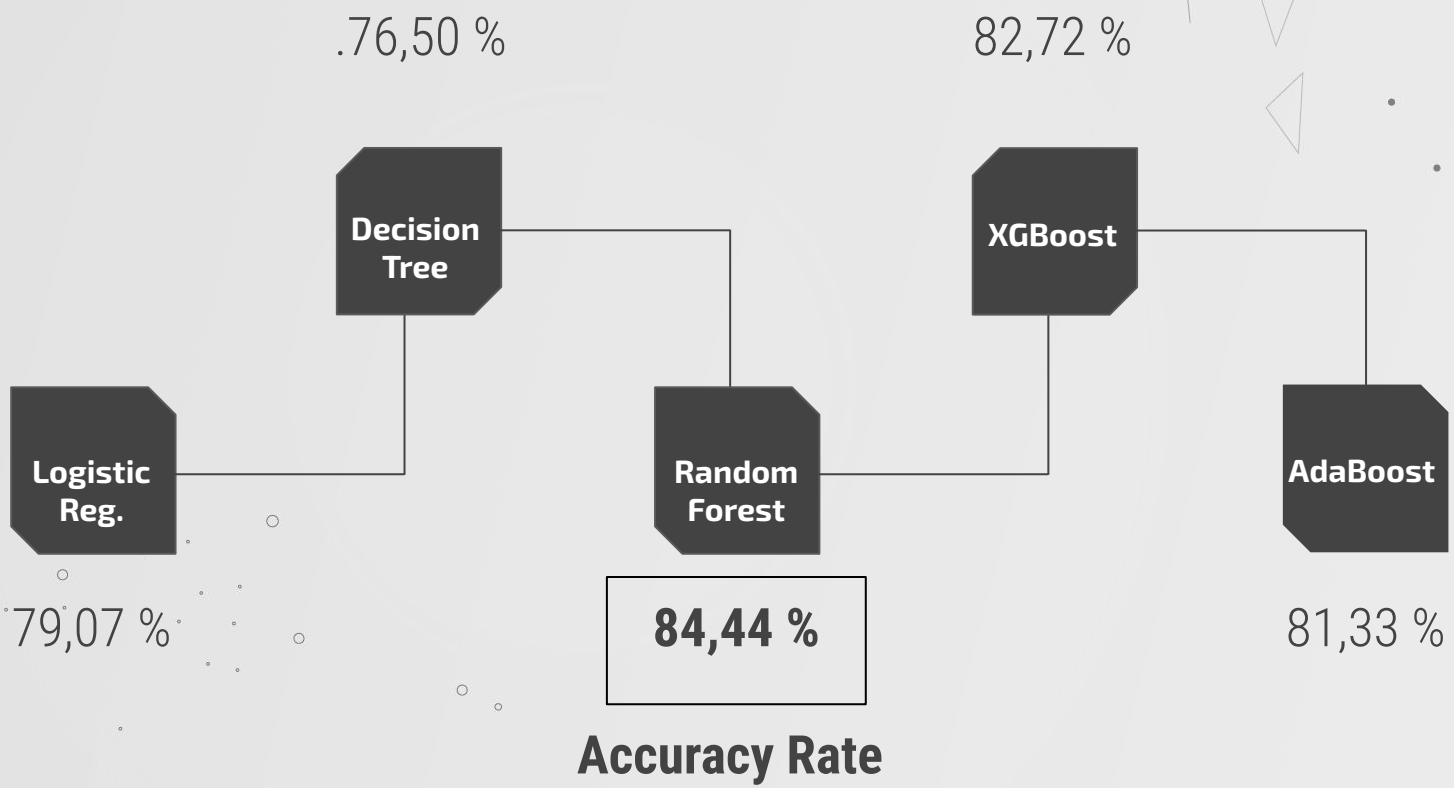
04

Results

Presenting results of the models



Bag-of-words



Accuracy Rate

Neural Network

Word2Vec:

- **Accuracy Rate around 80 %**
- **Need much time and energy**

05

Conclusion

Choosing best model

Recommendation

Bag-of-Words

- Accuracy Rate ~ 84 %
- Precision Rate ~ 76 %
- Most efficient model

Neural Network

- Small data set
- No need for Neural Networks

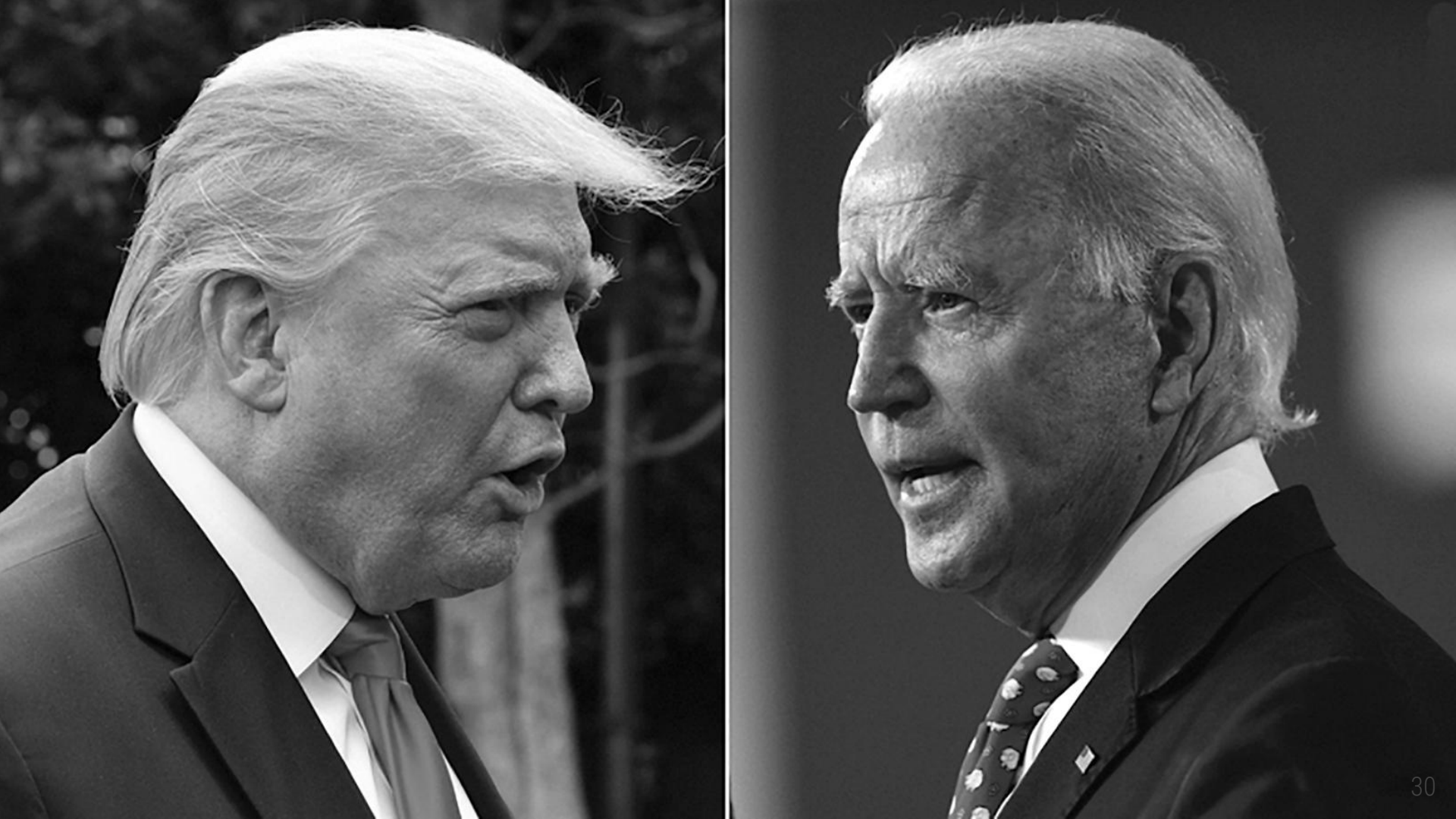
06

Outlook

Future Work

Future Work


Collect more data
Fake news detector for U.S. Presidential Election 2020
Improve Neural Network



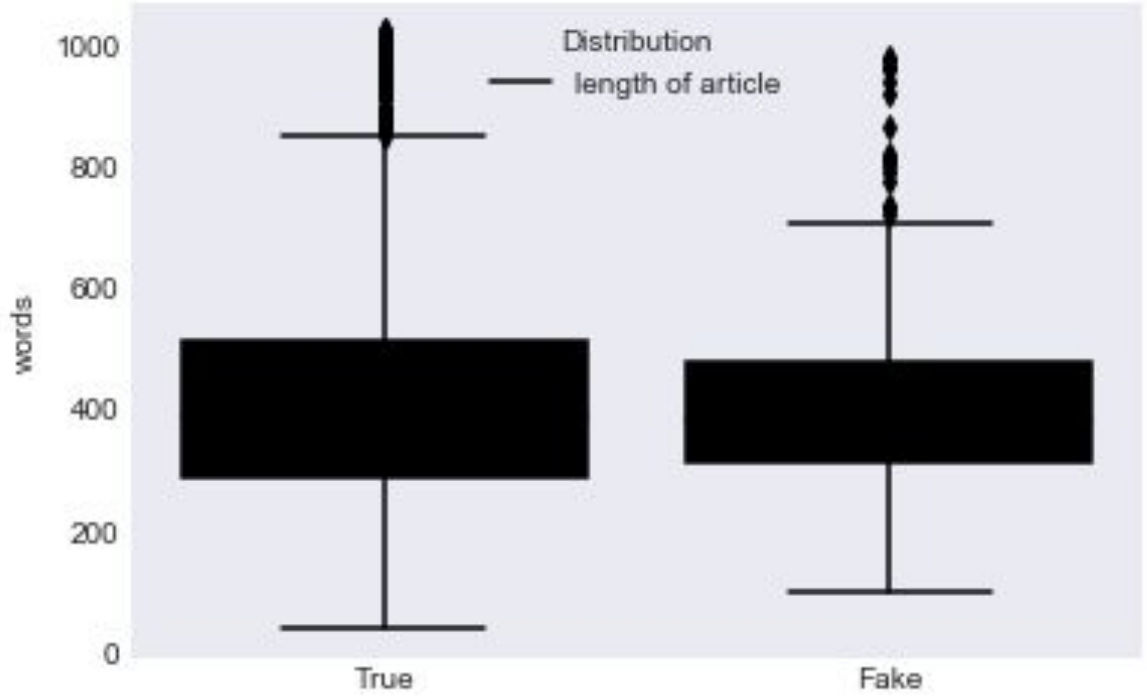
THANKS

Does anyone have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

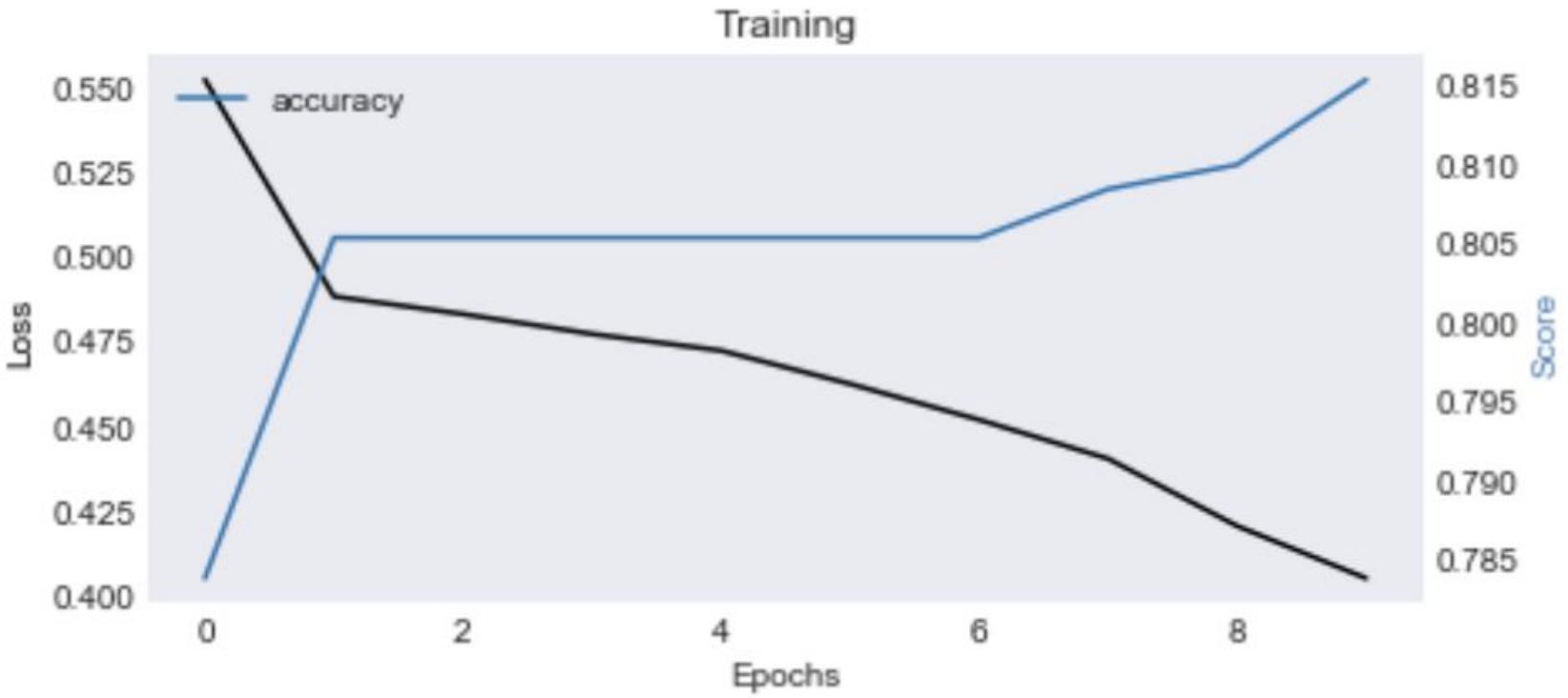
EDA: Box plot


Distribution
length of article
1000
800
600
400
200
0
words
True
Fake

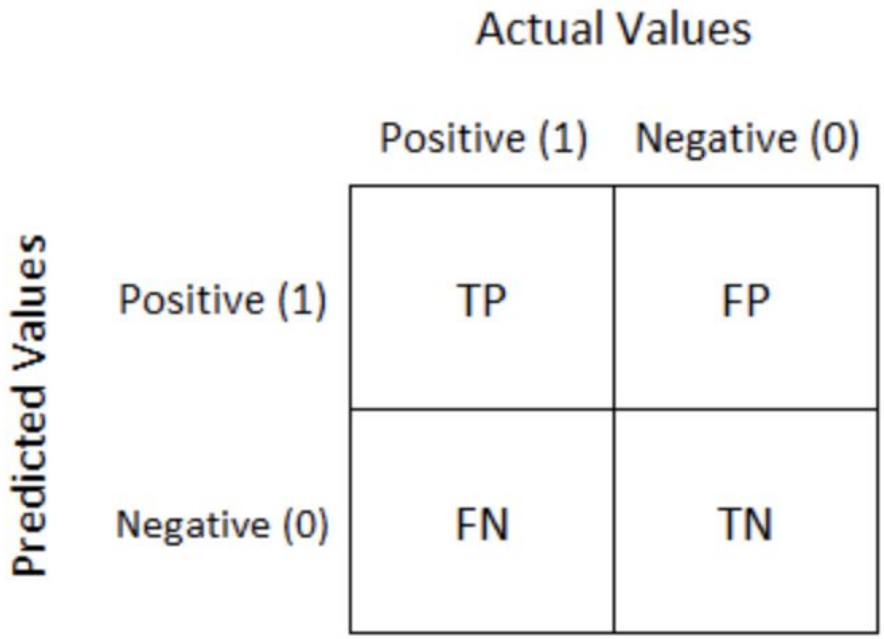
Bag-of-word: Results

	Accuracy	F1	Model	Precision	ROC-AUC	Recall
0	0.790773	0.624277	logreg	0.511041	0.794826	0.801980
1	0.765021	0.489510	tree	0.462555	0.676339	0.519802
2	0.844421	0.533762	forest	0.761468	0.687637	0.410891
3	0.827253	0.579634	XGB	0.613260	0.726807	0.549505
4	0.813305	0.571429	AdaBoost	0.568627	0.726855	0.574257

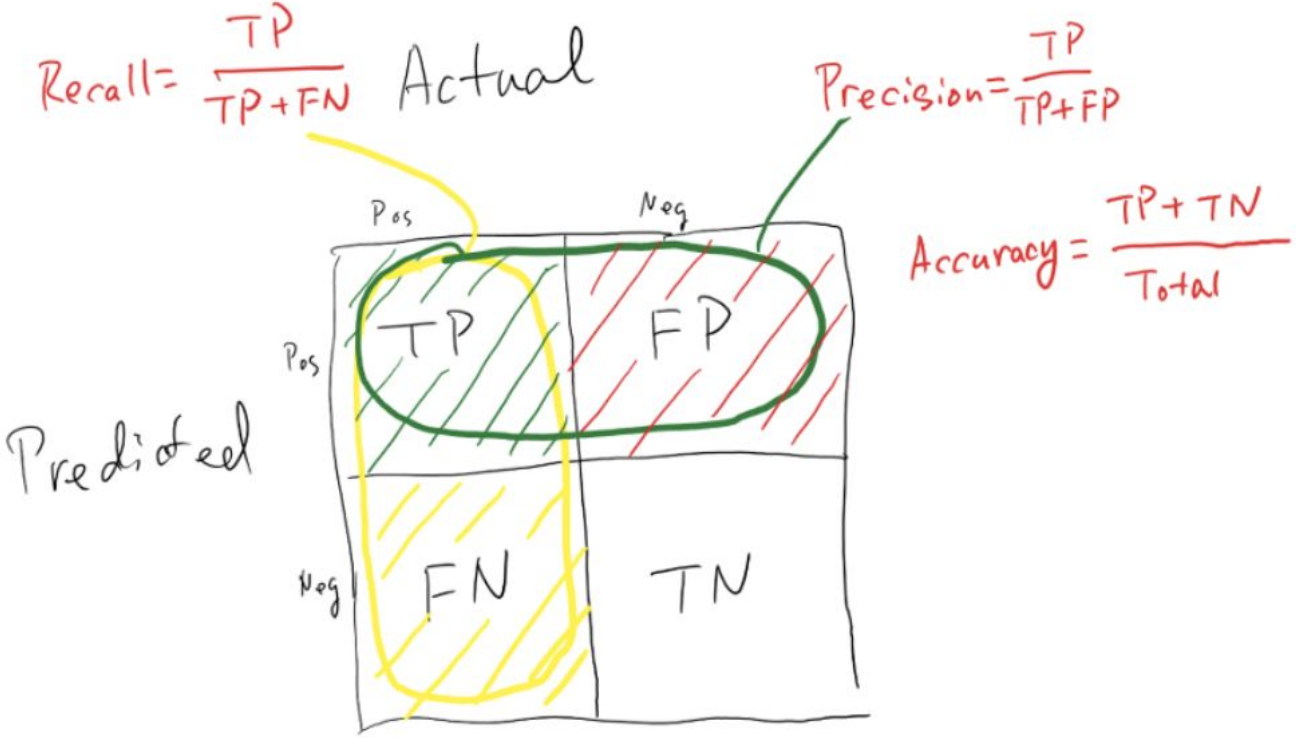
Word2Vec: Neural Network III



Confusion Matrix



Confusion Matrix II



RESOURCES

Did you like the resources on this template? Get them for free at our other websites.

VECTORS

- Technology background with gradient colors
- Blue 5g concept background
- Abstract landing pages with technology devices

PHOTOS

- Woman using smartphone with hologram
- Happy businesswoman looking at camera with holding pencil and diary
- Portrait of smiling man holding digital tablet looking at camera
- Smiling bearded man holding disposable coffee cup while opening door
- Portrait of pretty woman holding laptop looking at camera
- Motherboard with optical fiber