Cyber Security / Information Security：You are Fake News！

NTU ANTS lab 陳廷易*Leo*
臺灣好厲駭

Slide@GitHub:
tychen5/NLP_
FakeNewsDetection

Motivation

★ The effects of fake news flowed to the political world, causing the Internet fake news to fly during the 2016 US presidential election.

★ Trump is best known for making use of social networks to create public opinion, using exaggerated and ridiculous statements to create exposure.

★ Gartner trend report also estimated that the number of false news that most people will reach in 2020 will exceed real news.

Problem Description

★ In such an era when fake news is more than real news, we hope to analyze the difference between fake news and real news.

★ What kind of keywords are used in the real news and what characteristics are used in the fake news?

★ We want to find the difference between them so that readers will not be blindly deceived and influenced by fake news.

Agenda

★ Introduction

- ✓ background
- ✓ dataset

★ News Insight

- ✓ WordCloud & ScatterText
- ✓ Sentiment Analysis
- ✓ POS Tag Analysis

★ News Classification

★ Fake News Regression

Dataset

★ Using the relevant news during the US presidential election to analyze, and also the label published by signal media at Kaggle competition as ground truth (supervised **learning**).

★ In addition, using the fake news benchmark dataset published by UCSB for regression analysis.

★ Training news : 46589 documents (including 10% data for validation)

★ Testing news : 5200 documents

Eight Classes

★ **bias(扭曲)** — Traffic in political propaganda and gross distortions of fact.

★ **conspiracy(陰謀)** — Well-known promoters of kooky conspiracy theories.

★ **hate(仇恨)** — Promote racism, misogyny, homophobia, and other forms of discrimination.

★ **junk-sci(偽科學)** — Promote pseudoscience, metaphysics, naturalistic fallacies, and other scientifically dubious claims.

★ **satire(諷刺)** — Provide humorous commentary on current events in the form of fake news.

★ **state(受監督)** — Repressive states operating under government sanction.

★ **true(真新聞)** — Reliable.

★ **fake(假新聞)** — Fabricate stories out of whole cloth with the intent of pranking the public.

What is the Insight

WordCloud & ScatterText

Sentiment Analysis

POS Analysis

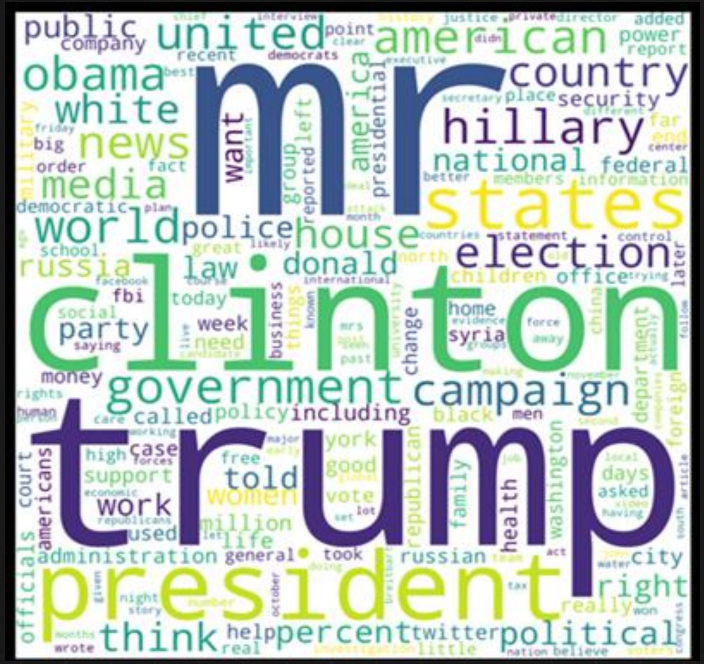
感謝組員陸艾眉提供

WordCloud and Scattertext

★ Filter out some common words (appearing more than 10,000 times)

★ mr, Trump, Clinton, president, America, Clinton, Hillary, etc.

★ In order to find out common words in each category.

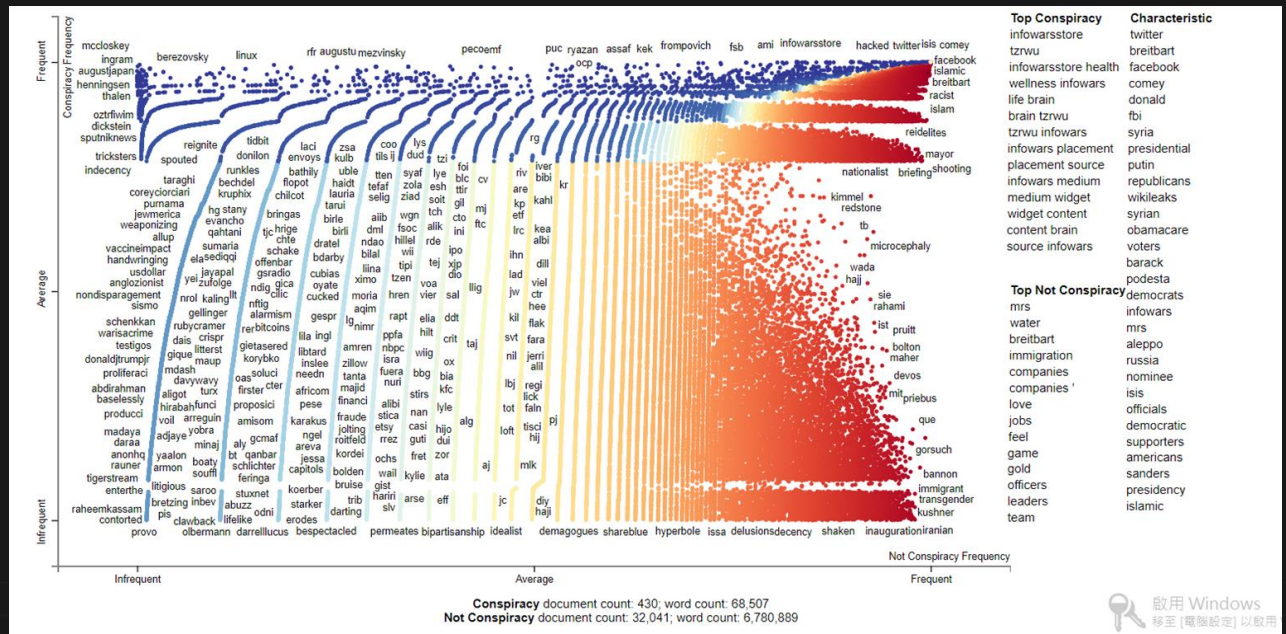


WordCloud and Scattertext

★In Conspiracy category, the **InfoWars** website and its contents are often found in this category.

★The top left corner words such as InfoWars store, health, brain, medium, etc.

★The bottom right corner such as immigration, water, Breitbart, etc.



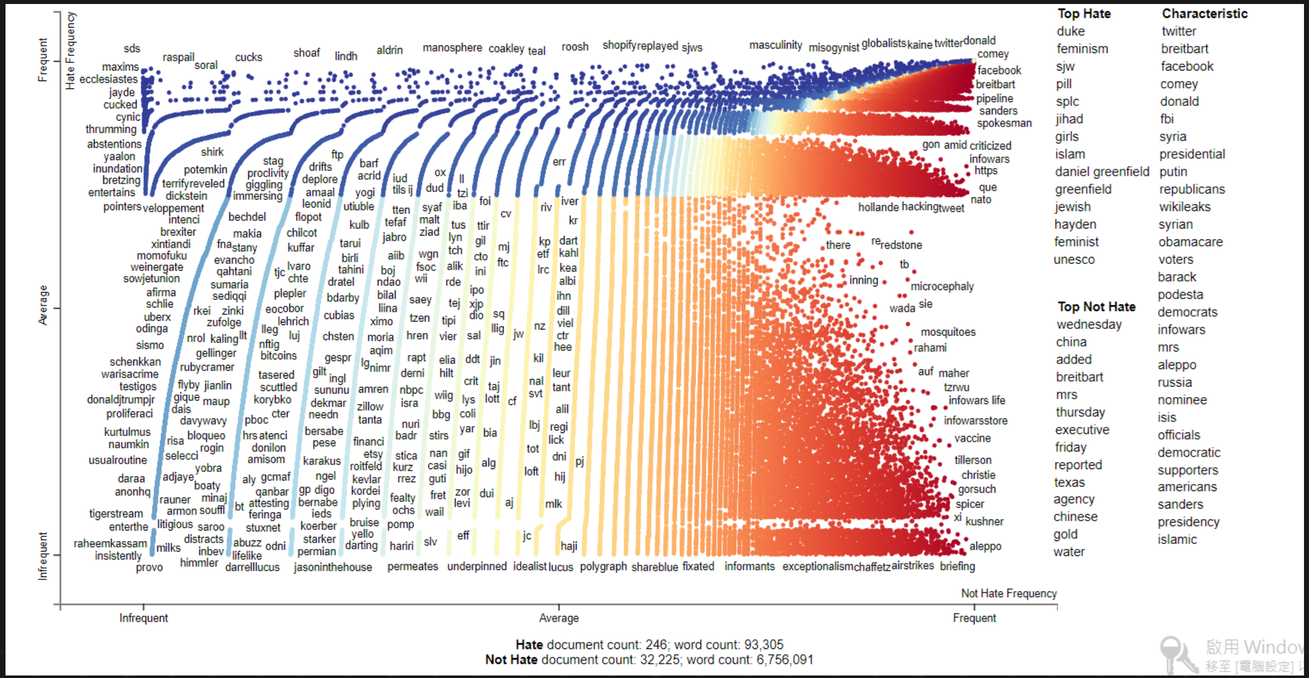
感謝組員陸艾眉提供

WordCloud and Scattertext

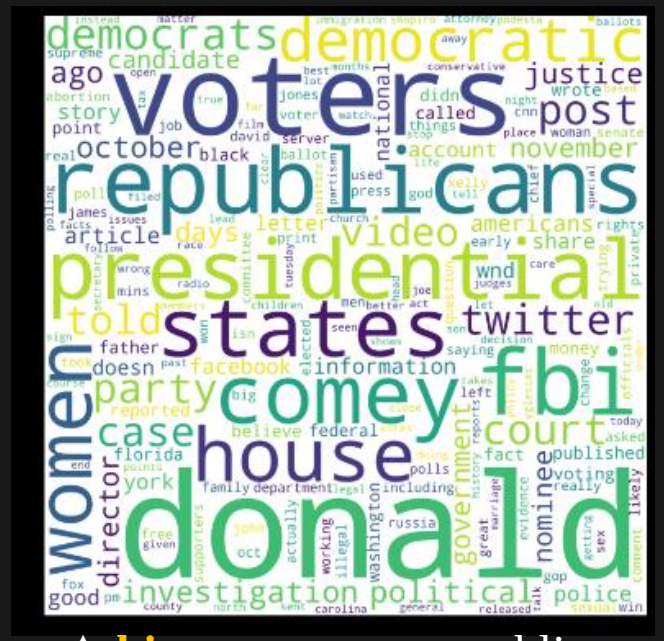
★In Hate category, most words were related to **gender** and **Islam**.

★The top left corner words such as female, islam, israel, etc.

★The bottom right corner words such as china, wednesday, gold, etc.

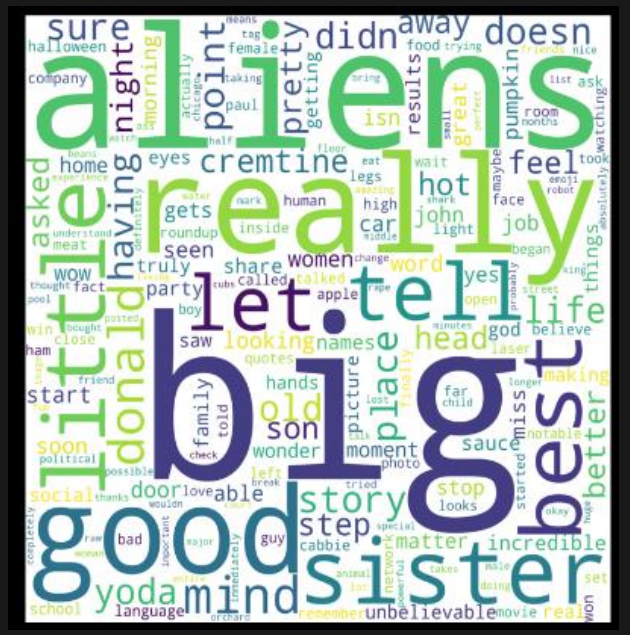


WordCloud and Scattertext



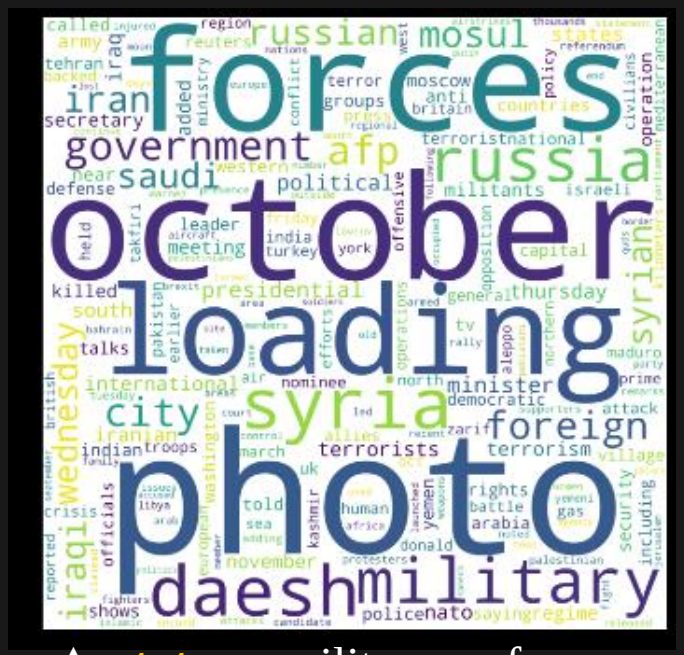
▲ **bias**: voters, republican, FBI, democratic, “Schapiro”

▲**junksci**: food, zika, health, natural, organic, vitamin

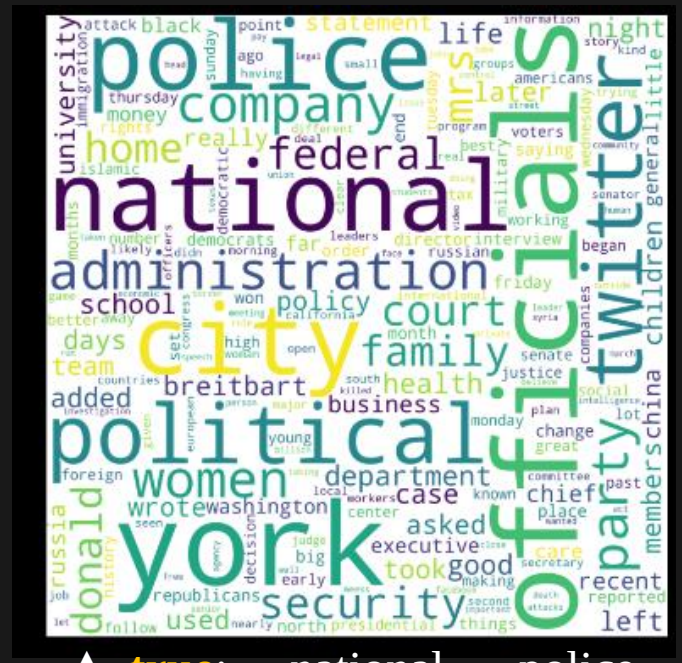


▲**satire**: aliens, really, good, “yoda”

WordCloud and Scattertext



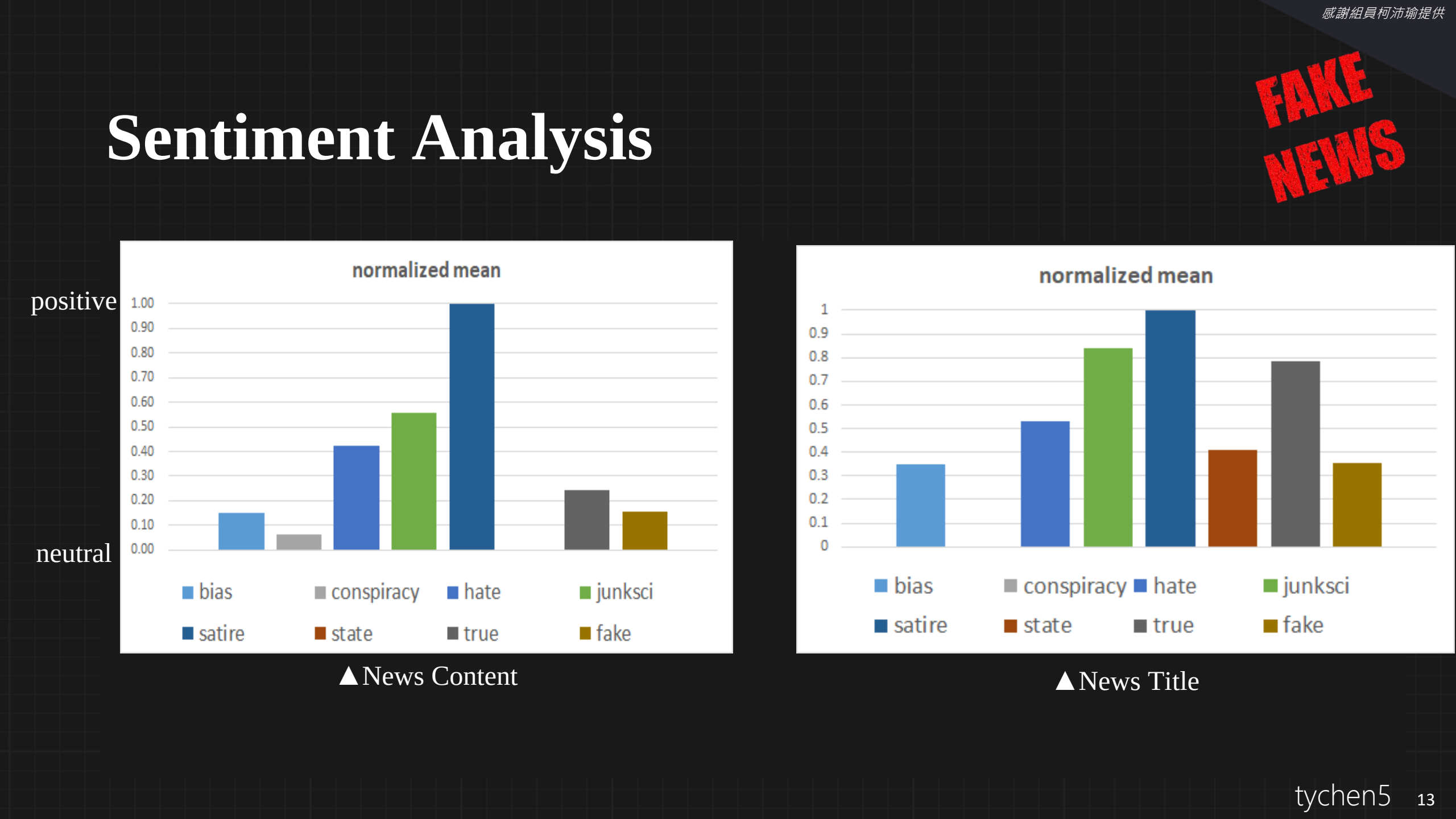
▲ **state**: military, forces, syria, “daesh”



▲ **true**: national, police, administration, “breitbart”

▲ **fake**: political, military, russia, anti, “podesta”

Sentiment Analysis



positive

neutral

▲News Content

▲News Title

Feature Selection (Statistics-based approach)

★ Compute feature utility

★ Reduce the number of terms used in text classification

★ Increase classification accuracy by eliminating noise features

★ Make training and applying a classifier more efficient

Feature Selection (Statistics-based approach)

★ Chi-Square statistic

>>10.83 : dependency

$$\chi^2(D,t,c) = \sum_{e_t \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{(N_{e_t e_c} - E_{e_t e_c})^2}{E_{e_t e_c}}$$

★ Log Likelihood Ratio (LLR)

more clear intuitive interpretable than chi-square

$$-2\log \frac{\left(n_{11}+n_{01}/N\right)^{n_{11}}\left(1-n_{11}+n_{01}/N\right)^{n_{10}} * \left(n_{11}+n_{01}/N\right)^{n_{01}}\left(1-n_{11}+n_{01}/N\right)^{n_{00}}}{\left(n_{11}/n_{11}+n_{10}\right)^{n_{11}}\left(1-n_{11}/n_{11}+n_{10}\right)^{n_{10}} * \left(n_{01}/n_{01}+n_{00}\right)^{n_{01}}\left(1-n_{01}/n_{01}+n_{00}\right)^{n_{00}}}$$

Feature Selection (Statistics-based approach)

★ Mutual Information

- ➢ Pointwise Mutual Information (PMI)

$$I(t,c) = \log_2 \frac{P(t \wedge c)}{P(t)P(c)}$$

  # The degree of uncertainty reduction of a class after a known term

  # Low-frequency terms will receive a higher score is irrational

  # Can be negative

- ➢ Expected Mutual Information (MI)

$$I(T,C) = \sum_{e_t \in \{1,0\}} \sum_{e_c \in \{1,0\}} P(e_t, e_c) \log \frac{P(e_t, e_c)}{P(e_t)P(e_c)}$$

Feature Selection (Statistics-based approach)

★Average TF-IDF scores

	d1-TF	d2-TF	d3-TF	d4-TF	IDF	d1-TFIDF	d2-TFIDF	d3-TFIDF	d4-TFIDF	AVG-TFIDF
cat	0	9	1	9	0.125	0	1.125	0.125	1.125	0.7916667
city	0	7	2	6	0.125	0	0.875	0.25	0.75	0.625
family	8	6	2	9	0	0	0	0	0	0
Leo	0	9	1	3	0.125	0	1.125	0.125	0.375	0.5416667

➢TF-IDF

$$tf_{idf_i} = tf_i * idf_i = \frac{n_{i,j}}{\sum_k n_{k,j}} * \log \frac{|D|}{|\{j: t_i \in d_j\}|}$$

High: (*t* occurs many times in *d*) and (appears within a small number of documents)

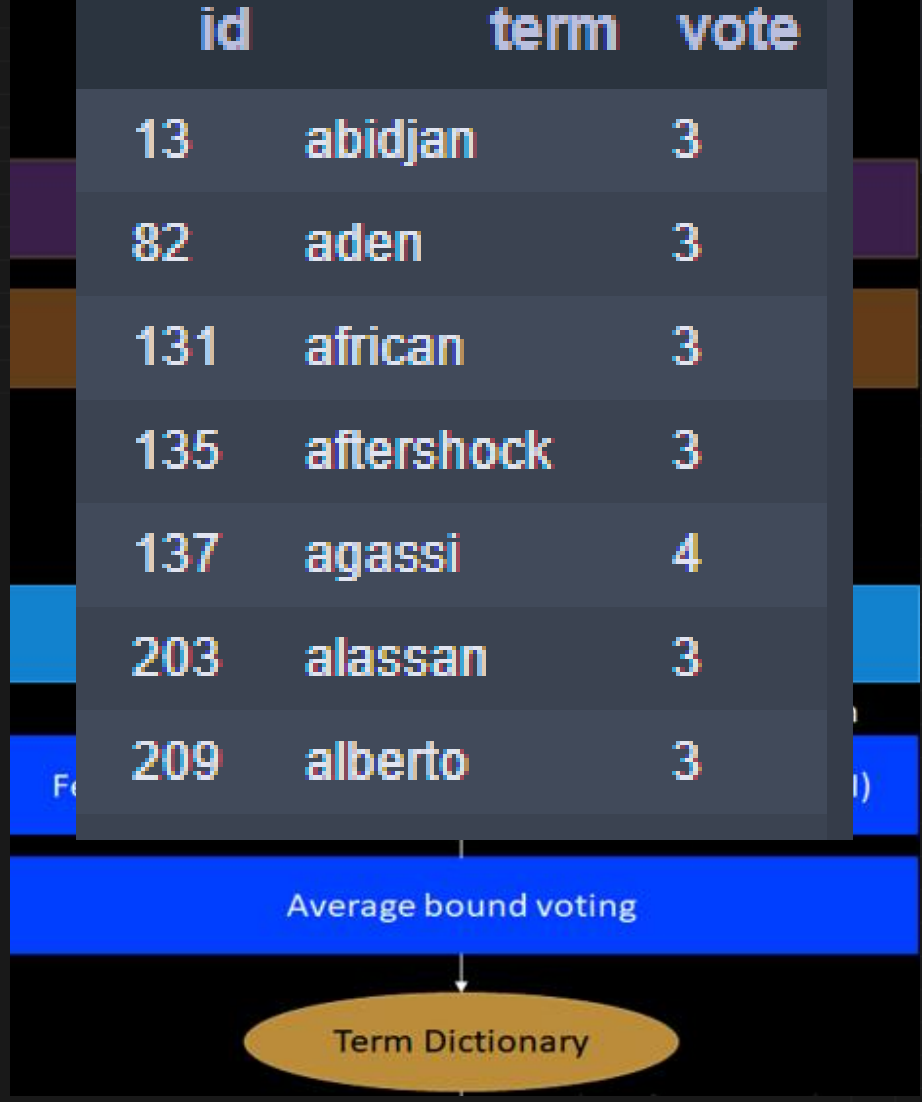
Low: (*t* is a rare term in *d*) and (occurs in virtually all documents in the collection)

Part-Of-Speech Tag Analysis

Tag	Meaning	English Examples
ADJ	**adjective**	**new, good, high, special, big, local**
ADP	adposition	on, of, at, with, by, into, under
ADV	**adverb**	**really, already, still, early, now**
CONJ	conjunction	and, or, but, if, while, although
DET	determiner, article	the, a, some, most, every, no, which
NOUN	**noun**	**year, state, costs, time, president**
NUM	numeral	twenty-four, fourth, 1991, 14:24
PRT	particle	at, on, out, over per, that, up, with
PRON	pronoun	he, their, her, its, my, I, us
VERB	**verb**	**is, say, told, given, playing, would**
.	punctuation marks	. , ; !
X	other	ersatz, esprit, dunno, gr8

POS processing

id	term	avg_tfidf	score_chi	score_llr	score_emi	vote
1	abandon	0.005374	3.846051e+00	1.659552	2.913791e-03	0
2	abc	0.015857	1.555876e+00	0.899990	1.580176e-03	0
3	abcnews.com	0.033696	6.544514e-01	0.204845	3.596454e-04	0
4	abdallah	0.006597	5.290393e-01	0.274067	4.811811e-04	0
5	abdel	0.003924	7.944005e-01	0.411467	7.224236e-04	0
6	abdomin	0.010949	5.290393e-01	0.274067	4.811811e-04	0
7	abduct	0.016260	3.668540e-01	0.160386	2.815866e-04	0
8	abdul	0.000851	9.267195e-01	0.464110	8.148566e-04	0



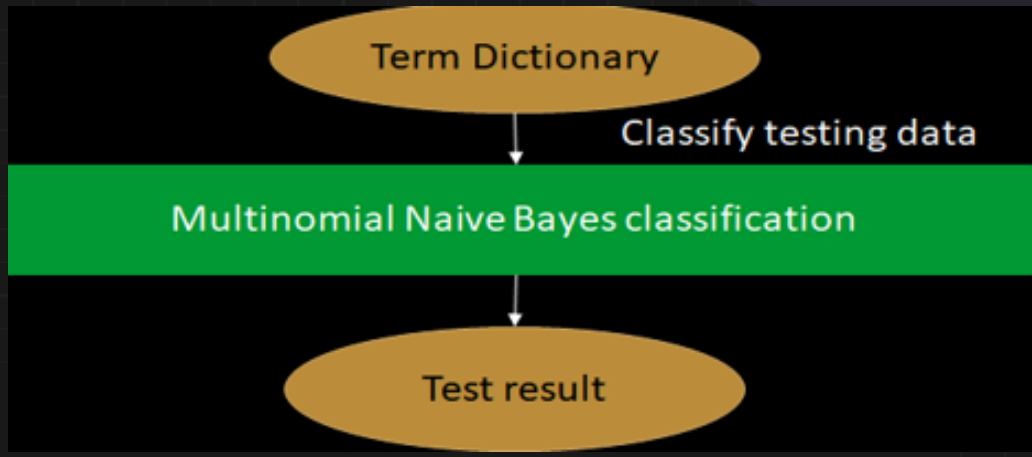
Feature Selection：

★ Chi-Square / Likelihood Ratio / Mutual Information / AVG TF-IDF

Voting：

★ Calculate each method's average score as criteria of a vote

POS processing



Testing：

★ Multinomial Naïve Bayes Classification

※Naïve Bayes：Efficiency & Effectiveness baseline classifier

- Multinomial model $c_{map} = \arg\max_{c \in C}[\log P(c) + \sum_{1 \le k \le n_d} \log P(X = t_k \mid c)]$
- Bernoulli model $c_{map} = \arg\max_{c \in C}[\log P(c) + \sum_{1 \le i \le M} \log P(U_i = e_i \mid c)]$

Classification results (POS)

Multinomial NB classifier：

★ w/o feature selection testing accuracy

- NOUN≒VERB≒ADJ≒ADV ≒ 0.47

★ with feature selection testing accuracy

- NOUN≒VERB ≒ 0.59
- ADJ≒ADV ≒ 0.58

What kind of news

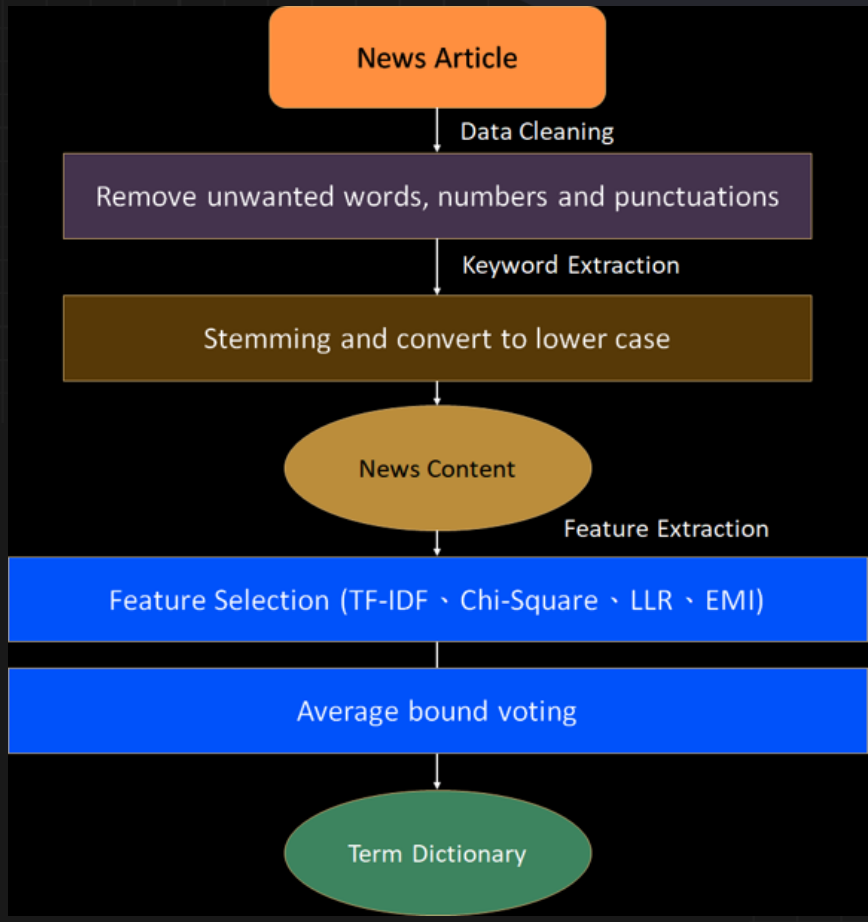
1000 / 7000 term dictionary

Multinomial Naïve Bayes

Support Vector Machine

Random Forest

Data Preprocessing



Data Cleaning：

★ Convert to lower case

★ Porter **stemming**

★ Filter out numbers, symbols and stop words

Feature Selection：

★ Chi-Square / Likelihood Ratio / Mutual Information / AVG TF-IDF

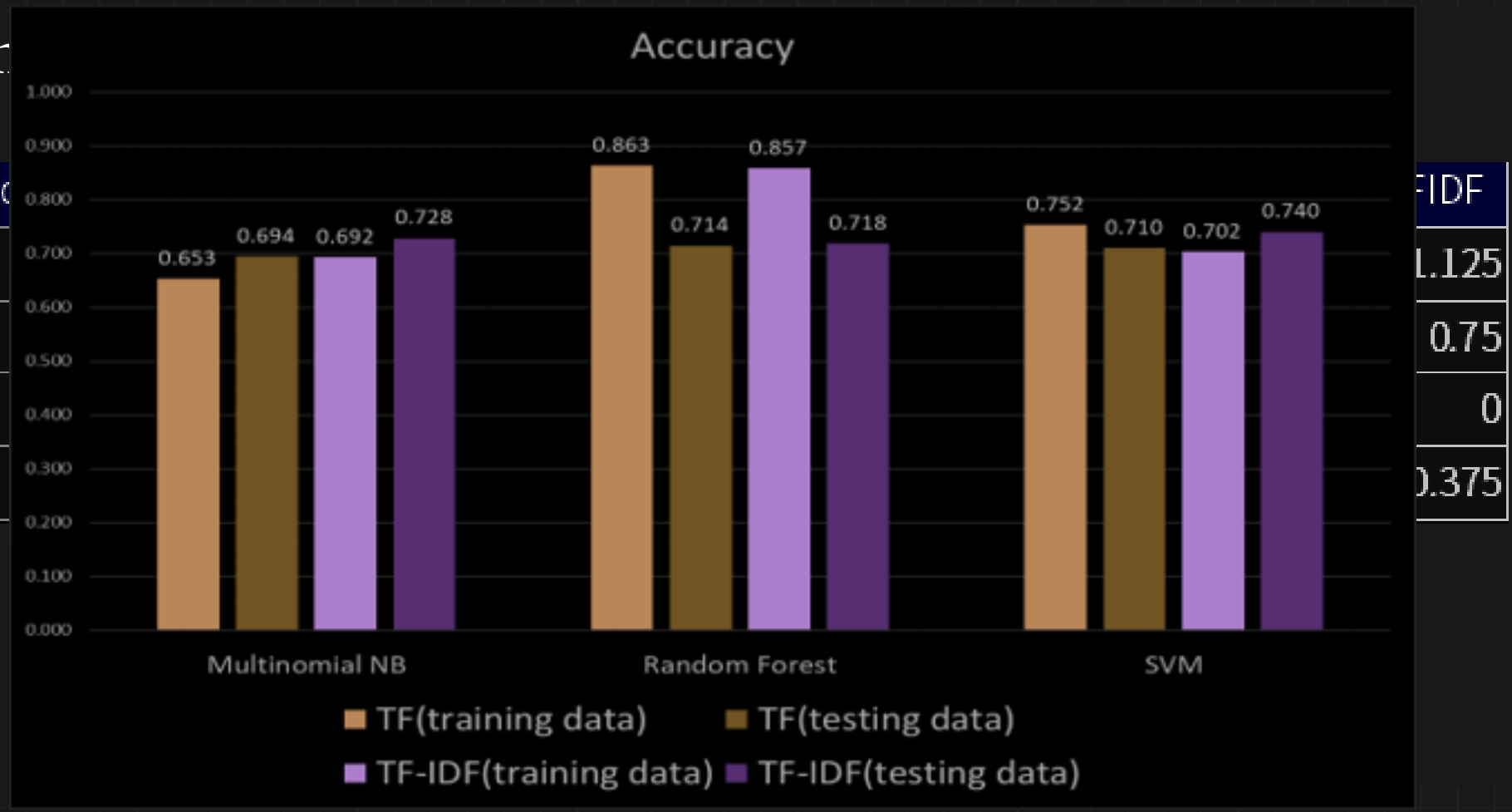
Voting Dictionary：

```
"And Yugoslav authorities are planning the arrest of eleven coal miners \nand two opposition politicians on suspicion of sabotage, tha
t's in \nconnection with strike action against President Slobodan Milosevic. \nYou are listening to BBC news for The World."
```

```
yugoslav author plan arrest eleven coal miner two opposit politician suspicion sabotag connect strike action presid slobodan milosev listen
bbc news world
```

★ Pick terms that have over one or two votes

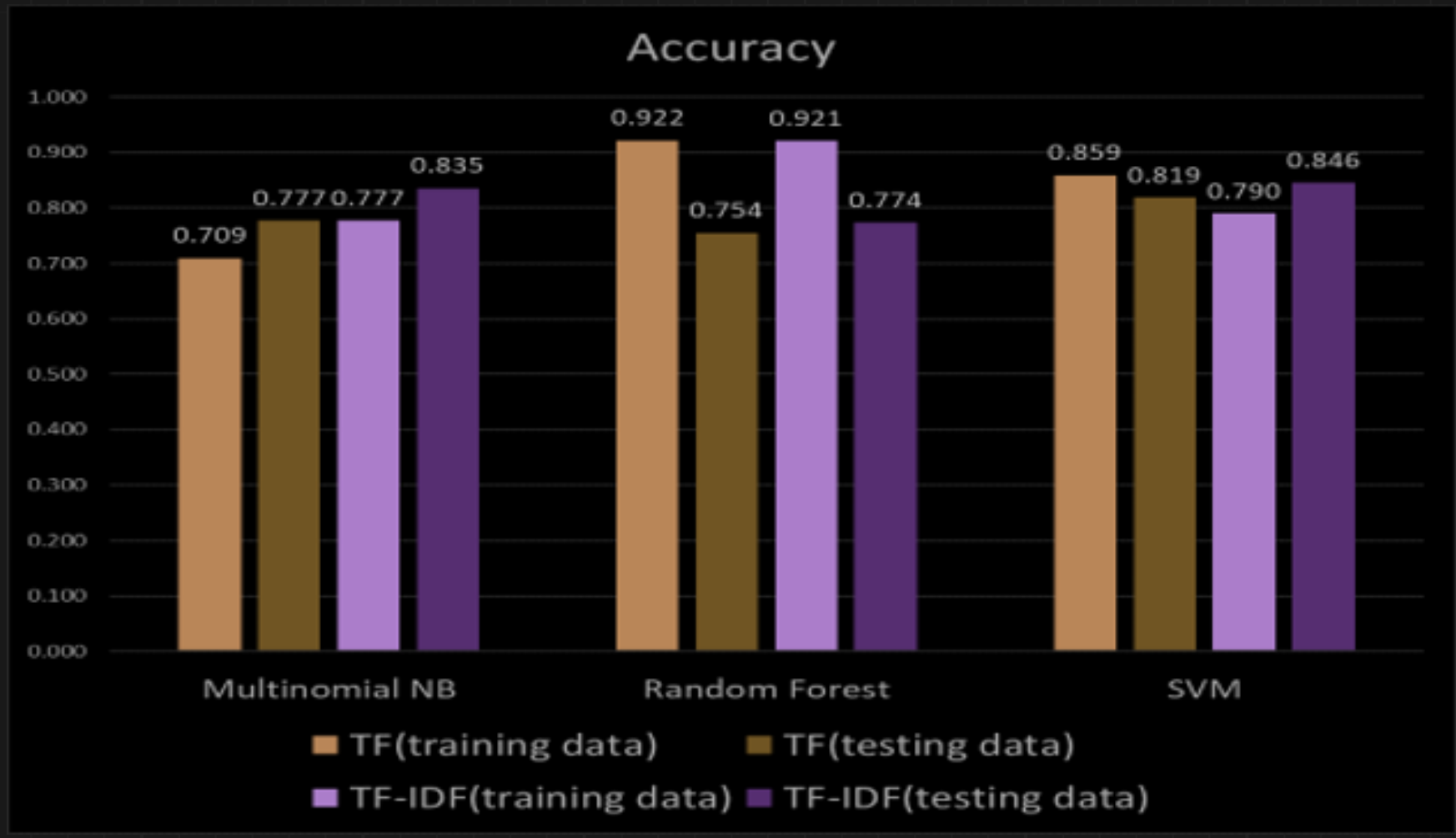
Classification Result (1000 term dictionary)



★ Trar

								FIDF
cat								1.125
city								0.75
family								0
Leo								0.375

Classification Result (7000 term dictionary)



How fake is the news

Convolutional Neural Network

Recurrent Neural Network

Text Regression

MSE / MAE

Six Levels

★ **True** — score equals to 1

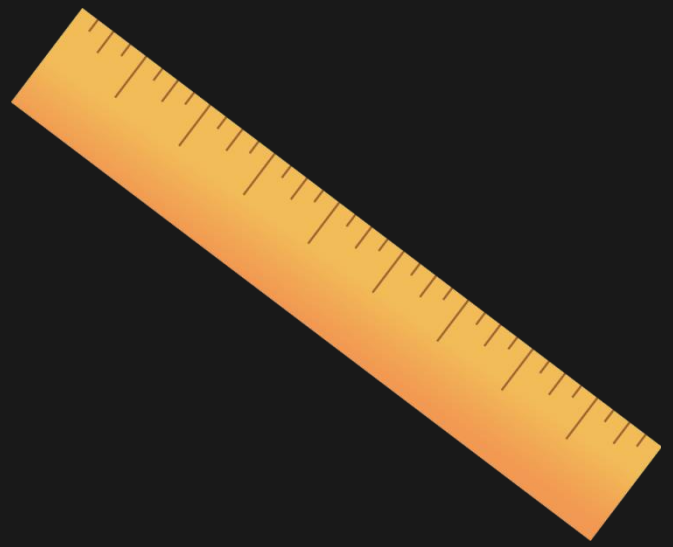
★ **Mostly-true** — score equals to 0.8

★ **Half-true** — score equals to 0.5

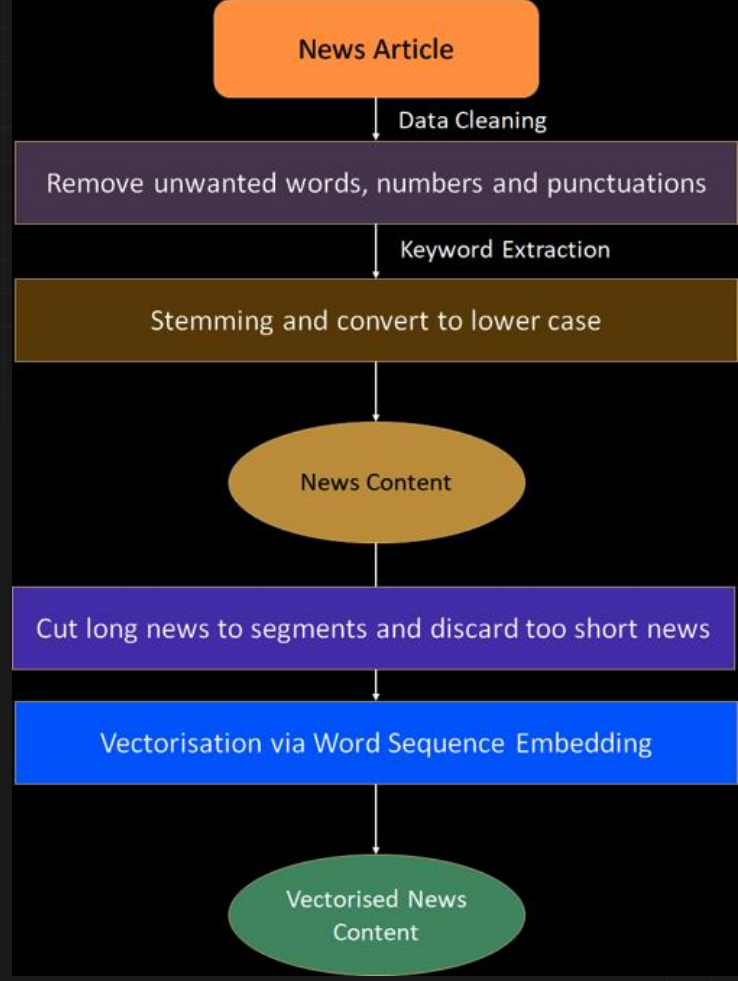
★ **Barely-true** — score equals to 0.2

★ **Fake** — score equals to 0.1

★ **Pants-on-Fire** — score equals to 0



Text Preprocessing



★ Preprocess same as categorical classification

★ Cut news longer than 237 terms to segments

- ✓ average number of terms in our dataset

★ Discard news shorter than 9 terms

- ✓ first knee point in our dataset

★ Use word embedding to vectorize news content (prediction-based approach)

- ✓ word embedding model will train with the whole regression model

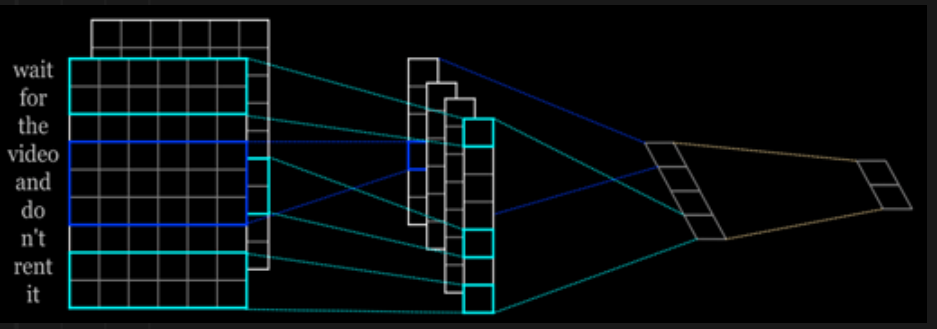
Model Design for Text Regression

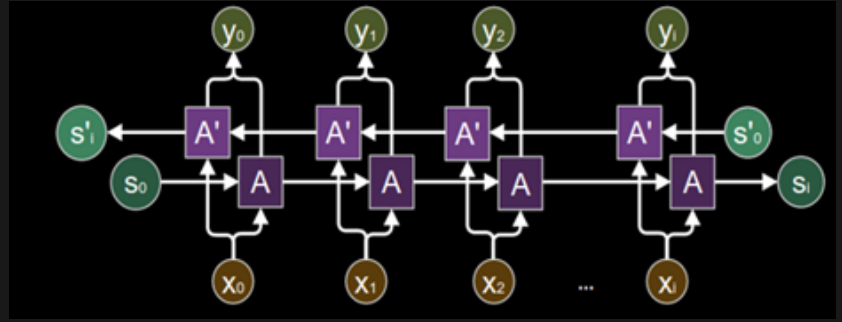
★ To capture key word pattern and retain temporal sequence

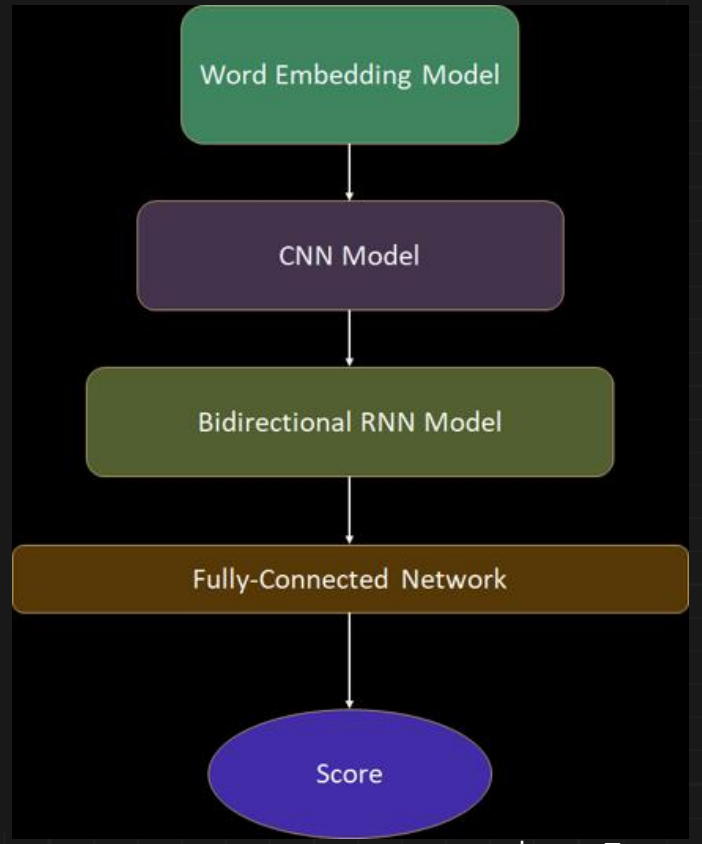
✓ dilated causal convolution (in one-dimension)

★ To remember word sequences in two ways

✓ bidirectional LSTM & GRU







Evaluation Metric

★ Mean Square Error = 0.067

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

★ Mean Absolute Error = 0.1

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

★ Binary classification accuracy = 0.945

✓ set score=0.5 as classifier threshold

System Demo

FAKE NEWS

Rick Perry: 'Historic record highs' (from terrorists states being appreh border

By Lauren Carroll on Sunday, August 3rd, 2014 at 4:05 p.m.

Texas Gov. Rick Perry appeared on CNN's "State of the Union" on Aug. 3, 2

Texas Gov. Rick Perry says the border needs more p number of illegal immigrants coming from terrorist match Perry's rhetoric.

With Congress scrambling to address illegal immigration before t summer recess this past Friday, it's no surprise that the issue was the Aug. 3, 2014, talk shows.

Perry, who is considering a 2016 presidential run, appeared on CNN's *State of the*


```
test_X = pad_sequences(test_X, maxlen=max_len, padding= 'post' )
test_ID = np.array(test_ID)
# test_X.shape , test_ID.shape

ans1 = model.predict(test_X)
ans2 = model2.predict(test_X)

ans = (ans1 + ans2)/2
ans = np.squeeze(ans)

ataFrame(data={'id':[int(test_ID)],'score':[float(ans)]})

ataFrame(data={'id':test_ID,'score':list(ans)})
roupby('id').mean().reset_index()
f.score.values)

7

got you! Lier,Lier,Pants on Fire!!')

!You are fake news!')

    print(' ','Well..It\'s mostly fake though..')
elif ans <=0.6:
    print(' ','Hm~~It\'s half-true and half-fake...')
elif ans <=0.9:
    print(' ','Um~It\'s mostly true^^')
else:
    print(' ','It\'s a true story~~!!')

 I got you! Lier,Lier,Pants-on-Fire!!
```

Conclusion

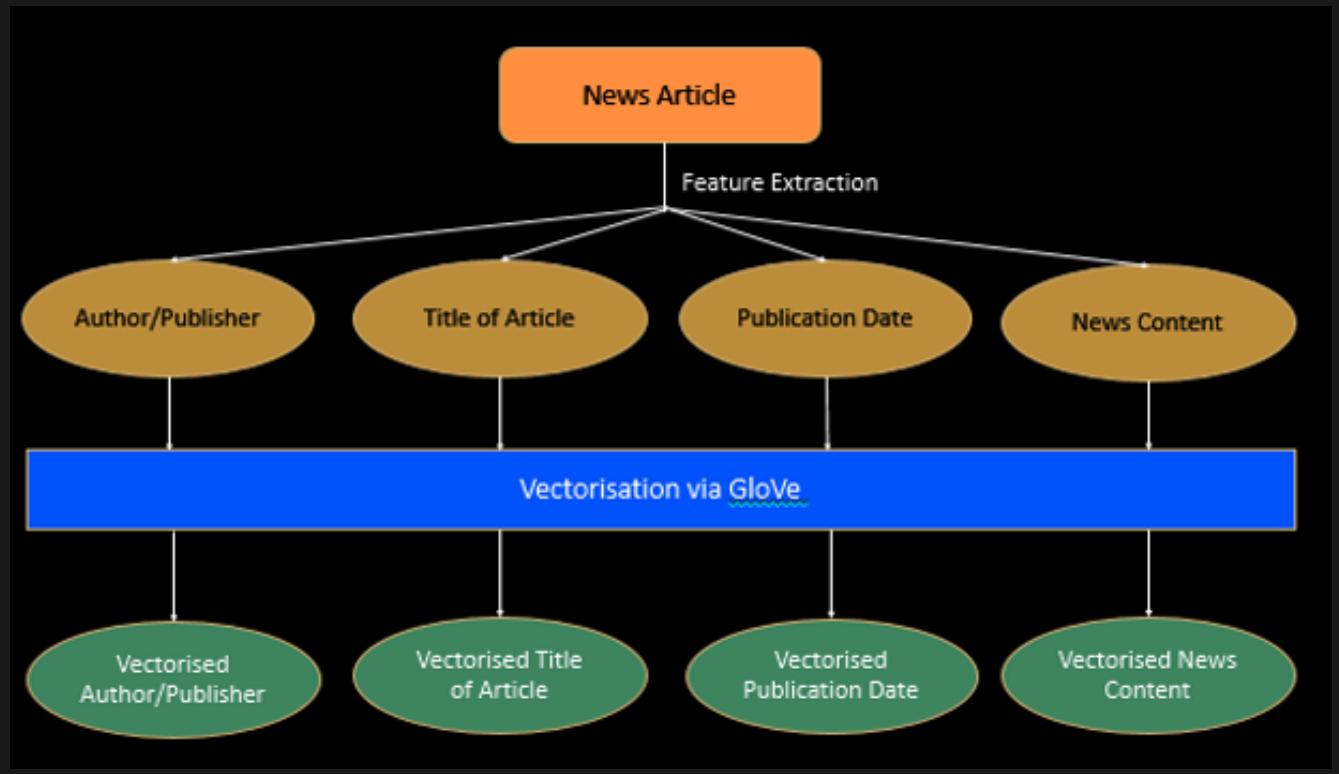
★ Define **FAKE** News ?

★ Domain specific task

★ Unite with other metadata

More than Content…

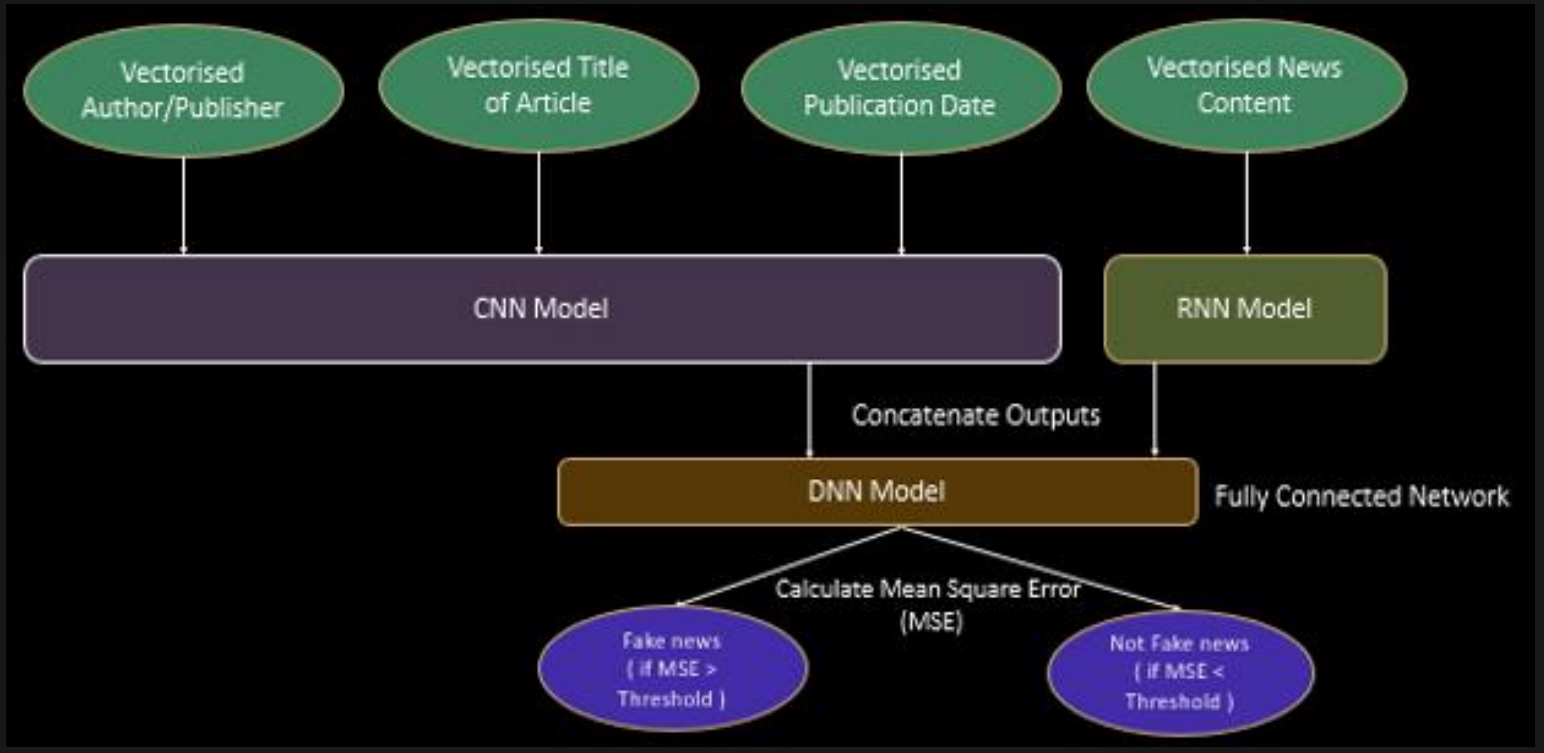
★ Author/Publisher, Title, Date



★ DEMO：https://ppt.cc/fbZZNx

More than Content…

★CNN: Author/Publisher, Title, Date ; RNN: Content



★ DEMO：https://ppt.cc/fbZZNx

Thank You！
& Crack the News！

NTU ANTSlab 陳廷易Leo

Related work & Reference

- ★ A novel text mining approach based on TF-IDF and Support Vector Machine for news classification https://ieeexplore.ieee.org/abstract/document/7569223
- ★ TEXT CLASSIFICATION USING NAÏVE BAYES, VSM AND POS TAGGER https://pdfs.semanticscholar.org/43d0/0d394ff76c0a5426c37fe072038ac7ec7627.pdf
- ★ Text categorization with Support Vector Machines: Learning with many relevant features https://link.springer.com/content/pdf/10.1007%2FBFb0026683.pdf
- ★ Unsupervised Content-Based Identification of Fake News Articles with Tensor Decomposition Ensembles: http://snap.stanford.edu/mis2/files/MIS2_paper_2.pdf

Q & A

GitHub: tychen5

LinkedIn: tychen5

Best Project Award
& Detail Report