

Assignment 3 Analysis



United We Stand:
Sunny Patel (40075684)
Zlata Korotkova (40051921)

Initial Dataset

Training Set:

- 399 instances
- Unbalanced
- Symbols

Total true tweets: 247

Total fake tweets: 152

ained hotel room no NA
188.50 What youâ€SGCId
of congress prophylatic
OVID-19 "Chinese coron
wonâ€SGCIt hear from th
ratic hoax to hurt Trum

ALERT!!!!!! The corona virus can be spread
COVID-19 health advice⚠️ <https://t.co/XsSA>
⚠️Chinese doctors say autopsies of #corona
.@NationalNurses President Deborah Burger
WOW. A man in Incheon, west of Seoul - wh
So the CORONA virus stopping y'all from e
Miss corona u have another job to do 🥰🥰
ALERT!!!!!! The corona virus can be spread

Initial Dataset

Test Set:

- 54 instances
- Symbols

ws and do res
@raphbosano
ed every week

os Here are information on all 33 c
ere then... <https://t.co/GQoW8ob3ti>
canâ€SGCIt beat this guy ffs <https://t.co/GQoW8ob3ti>
to get tested for the #coronavirus.

Models' Vocabulary

NB-BOW-OV:

- 4314 words
- Worse performance

```
Accuracy: 0.6666666666666666
Precision: 0.6666666666666666 0.6666666666666666
Recall: 0.9090909090909091 0.2857142857142857
F1: 0.7692307692307692 0.4
```

$$H_{NB} = \underset{H_i}{\operatorname{argmax}} P(H_i) \times \prod_{j=1}^n P(a_j | H_i)$$

NB-BOW-FV:

- 1193 words
- No unnecessary words (overfitting)

```
Accuracy: 0.7592592592592593
Precision: 0.7380952380952381 0.8333333333333334
Recall: 0.9393939393939394 0.47619047619047616
F1: 0.8266666666666667 0.6060606060606061
```

$$P(w_i | c_j) = \frac{(\text{frequency of } w_i \text{ in } c_j) + \cancel{1} \times \mathbf{0.01}}{\text{total number of words in } c_j + \text{size of vocabulary} \times \mathbf{0.01}}$$

```
P(true) = true_tweets / total_tweets
P(false) = false_tweets / total_tweets
```

Results of All 3 Models

NB-BOW-OV:

```
Accuracy: 0.6666666666666666
Precision: 0.6666666666666666 0.6666666666666666
Recall: 0.9090909090909091 0.2857142857142857
F1: 0.7692307692307692 0.4
```

NB-BOW-FV:

```
Accuracy: 0.7592592592592593
Precision: 0.7380952380952381 0.8333333333333334
Recall: 0.9393939393939394 0.47619047619047616
F1: 0.8266666666666667 0.6060606060606061
```

LSTM:

```
Accuracy: 0.7455
Precision: 0.7209 0.8333
Recall: 0.9394 0.4545
F1: 0.8158 0.5882
```

Results of All 3 Models

- Unbalanced training dataset
- Recall: we don't want to label fake tweets as factual
- Best performing model: NB-BOW-FV
- Worst performing model: NB-BOW-OV

Total true tweets: 247

Total fake tweets: 152

```
accuracy = totalCorrectPredictions / totalPredictions
precisionYes = correctYes / (correctYes + yesFP)
precisionNo = correctNo / (correctNo + noFP)
recallYes = correctYes / (correctYes + yesFN)
recallNo = correctNo / (correctNo + noFN)
f1Yes = 2 * precisionYes * recallYes / (precisionYes + recallYes)
f1No = 2 * precisionNo * recallNo / (precisionNo + recallNo)
```

Responsibilities

Sunny Patel:

- NB-BOW-OV
- NB-BOW-FV
- Trace files

Zlata Korotkova:

- Evaluation files
- LSTM
- Presentation

Thank You! :)