Fake News Detection

A Comparative Study of Computational Methods

Advanced Machine Learning
Academic Year 2023-2024

Nicolò Urbani 856213
Mattia Piazzalunga 851931

Fake or Real ?

Miracle! Pothole Turns Lifesaver After 'Dead' Haryana Man Comes to Life as Ambulance Hits Cavity

• Curated By: Sheen Kachroo • News18.com • Last Updated: JANUARY 13, 2024, 03:48 IST • Haryana, India

On the way, the ambulance hit a pothole hard and Brar's hand moved. Singh also checked his heartbeat and upon finding he brought the 80-year-old to a neighboring hospital. (Representative photo/PTI)

When the ambulance hit the pothole, Brar's grandson noticed his hand movement and sensed his heartbeat. He immediately asked the driver to head toward the neighbouring hospital. It was found that Brar was alive

Source: INDIAN NEWS, 2024- https://www.ndtv.com/india-news/dead-man-comes-alive

Fake News

"News that conveys or incorporates false, fabricated, or deliberately misleading information, or that is characterized as or accused of doing so"
- Oxford English Dictionary-

Social Media

Social Increase the diffusion of fake news

Absence of Verification

Just 23 % of people verify the Truthfulness news

Big Influence on Society

Influence public opinion, harm the reputation, and even pose a threat to national security.

Fake news

Source: EUROSTAT, 2021 - https://ec.europa.eu/eurostat/web

Project Purpose

The objective of this project is to compare and evaluate the performance of different Deep Learning and NLP techniques on various datasets and determine which technique is the most effective across different datasets for fake news classification

Analysis Fake News Language

Fake News use a specific language?

Preprocessing Impact

Does NLP pipeline improve the performance?

Compare different Classifier

Which model achieves the best performance?"

Augmentation Impact

Augmentation Techniques Improve the performance?

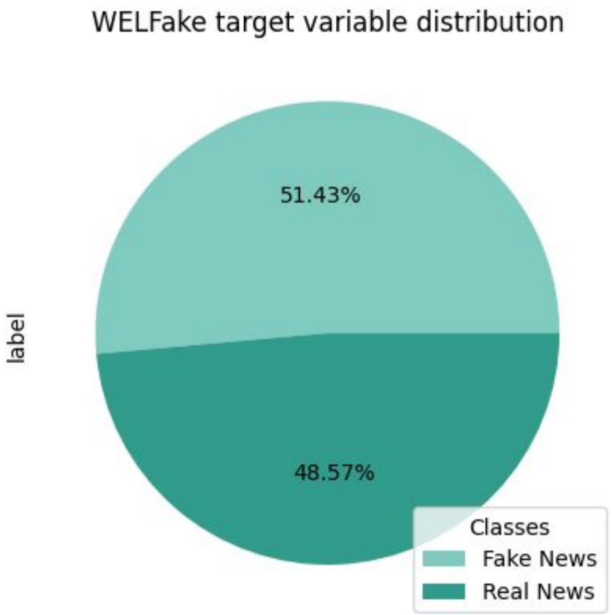
WELFake Dataset

The dataset **WELFake** contains **72.134 news** article from different sources: Kaggle, McIntire, Reuters, and BuzzFeed Political ecc. divided in:

- 35.028 ”real news”
- 37.106 ”fake news”

The dataset contains the following attributes:

- title: the title of a news article
- text: the text of the article; could be incomplete
- label: 1 for Real News, 0 for Fake News



Source: VERMA ET AL., 2021 - https://ieeexplore.ieee.org/document/9395133/

ISOT Dataset

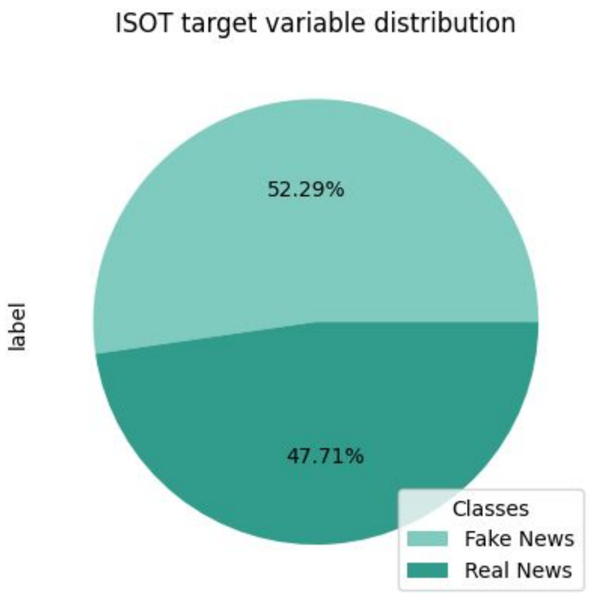
The dataset ISOT, produced by Canadian University of Victoria’s research team, contains **44.898 news** of year 2016 article divided in:

- 21417 ”real news” from Reuters.com
- 23481 ”fake news” from Kaggle.com

The news has been classified by Facebook and Polifact (an American Organization that works with Fact Checking)

The dataset contains the following attributes:

- Title: the title of a news article
- Text: the text of the article; could be incomplete
- label: 1 for Real News, 0 for Fake News



Source: AHMED ET AL., 2017 - https://link.springer.com/chapter/10.1007/978-3-319-69155-8_9

Dataset Sample - WELFake

FAKE NEWS

LAW ENFORCEMENT ON HIGH ALERT Following Threats Against Cops And Whites On 9-11By #BlackLivesMatter And #FYF911 Terrorists [VIDEO]

No comment is expected from Barack Obama Members of the #FYF911 or #FukYoFlag and #BlackLivesMatter movements called for the lynching and hanging of white people and cops. They encouraged others on a radio show Tuesday night to turn the tide and kill white people and cops to send a message about the killing of black people in America.One of the F***YoFlag organizers is called Sunshine. She has a radio blog show hosted from Texas called, Sunshine s F***ing Opinion Radio Show. A snapshot of her #FYF911 @LOLatWhiteFear Twitter page at 9:53 p.m. shows that she was urging supporters to Call now!! #fyf911......

0 Fake News
1 Real News

	title	text	label
0	LAW ENFORCEMENT ON HIGH ALERT Following Threat...	No comment is expected from Barack Obama Membe...	0
1	NaN	Did they post their votes for Hillary already?	0
2	UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MO...	Now, most of the demonstrators gathered last ...	0
3	Bobby Jindal, raised Hindu, uses story of Chri...	A dozen politically active pastors came here f...	1
4	SATAN 2: Russia unvelis an image of its terrif...	The RS-28 Sarmat missile, dubbed Satan 2, will...	0

ISOT - Dataset Sample

The ISOT dataset shares the same structure as the WELFake dataset, suggesting compatibility between the two for certain purposes

0 Fake News
1 Real News

	title	text	label
0	CAN YOU GUESS THE ONE THING Majority Of Bernie...	The function of socialism is to raise sufferin...	0
1	Congressional Black Caucus: Jeff Sessions Has...	The Congressional Black Caucus is a very influ...	0
2	Crybaby “Safe Space” Students Are Put On Notic...	This letter is quite possibly the most importa...	0
3	Macron assures Iran's Rouhani of France's comm...	PARIS (Reuters) - French President Emmanuel Ma...	1
4	France unveils labor reforms in first step to ...	PARIS (Reuters) - French President Emmanuel Ma...	1

Project Steps

1 Explorative Analysis

Problem Understanding and first analysis

2 Preprocessing

Data Cleaning and Preprocessing

3 Vectorization and Embedding

Use of Glove Embeddings

4 Classification Model

Comparison of different Model

5 Data Augmentation & Distillation

Increase the generalization of the models and reduce the models parameters

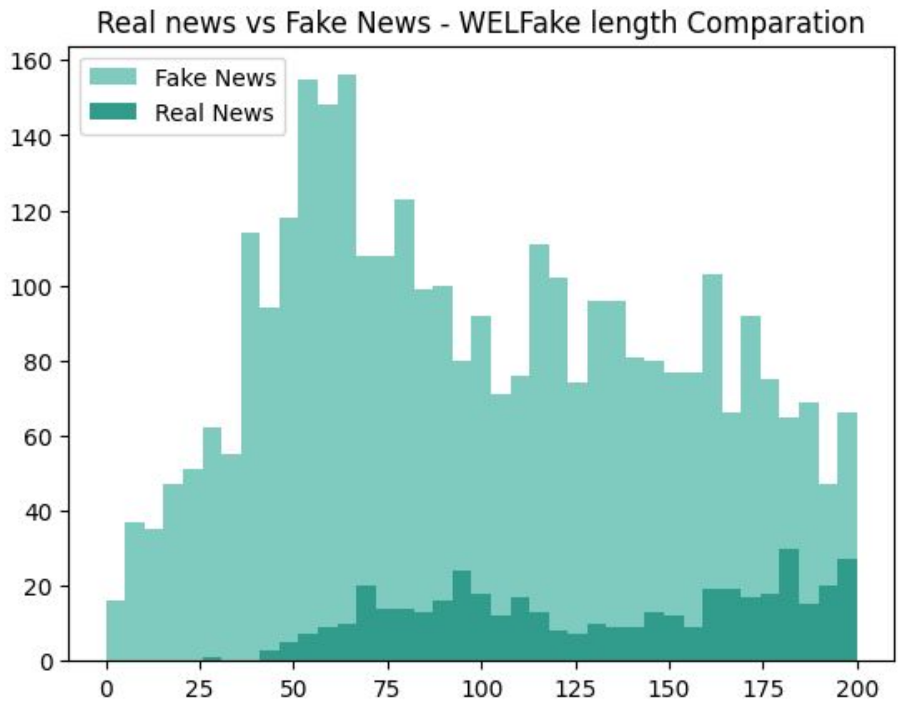
6 Analysis of Results

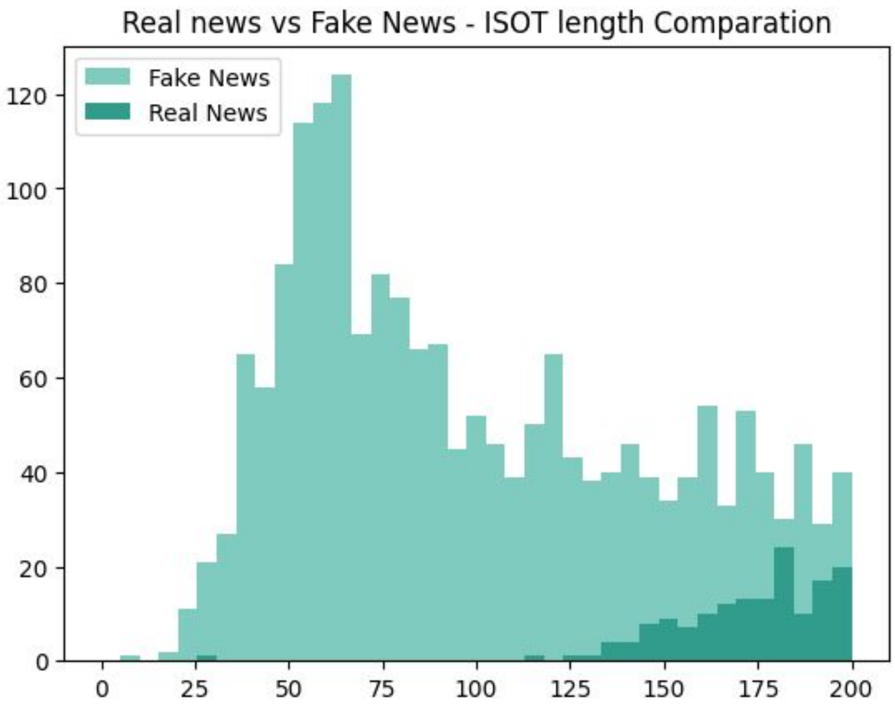
Analysis of performance and future development

1

Explorative Analysis

Fake vs Real Length Comparation





2

Preprocessing & Data Cleaning

Dataset Preprocessing

Remove NaN

Identify all the NaN in the dataset, this will be useful for preprocessing

1

Join Title & Text

The title and text fields will be considered jointly, they will be treated as an integrated unit

2

Remove Empty Text

After joining title and text are NaN the row will be deleted

3

Index Reset

The index of the dataset will be reset to avoid potential problems.

4

Dataset Preprocessing - Example

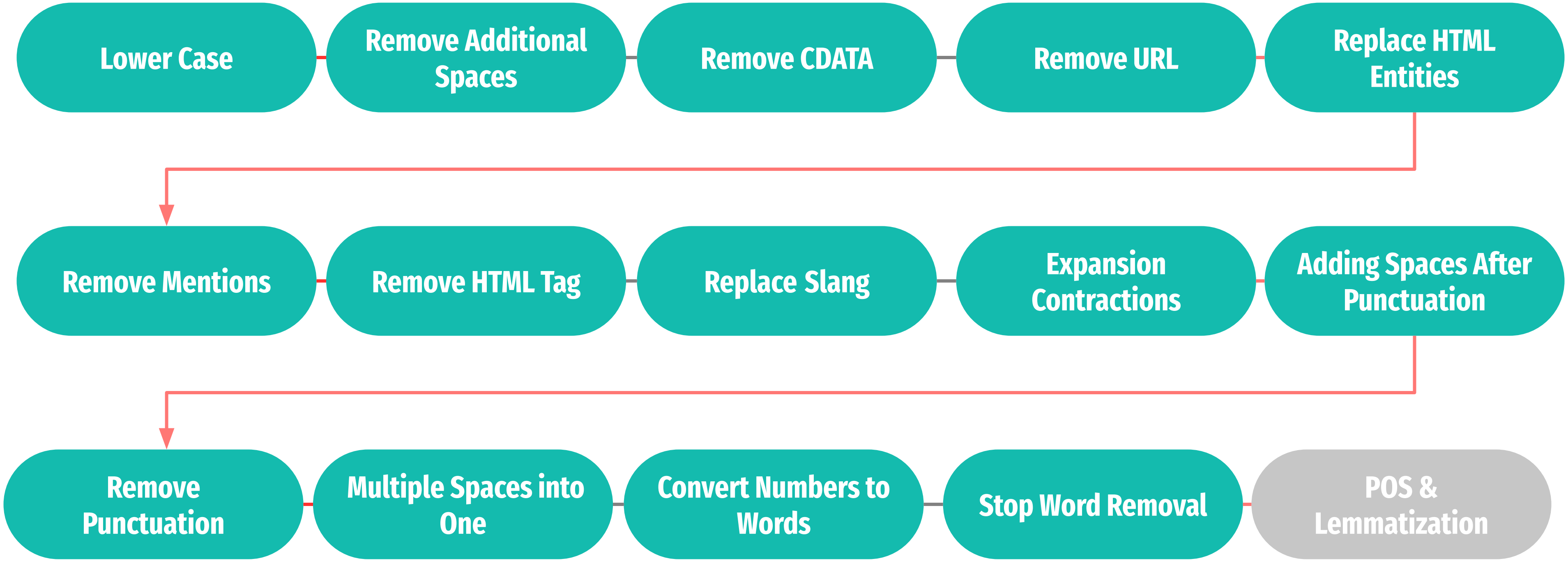
	title	text	Label
0	LAW ENFORCEMENT ON HIGH ALERT Following Threat...	No comment is expected from Barack Obama Member...	1
1	NaN	NaN	0
2	NaN	Did they post their votes for Hillary already?	0

DATASET PREPROCESSING

Title + Text

	text	Label
0	LAW ENFORCEMENT ON HIGH ALERT Following Threat... No comment is expected from Barack Obama Member...	1
1	Did they post their votes for Hillary already?	0

Text Cleaning



Text Cleaning - Example

BEFORE

LAW ENFORCEMENT ON HIGH ALERT Following Threats Against Cops And Whites On 9-11By #BlackLivesMatter And #FYF911 Terrorists [VIDEO]
No comment is expected from Barack Obama Members of the #FYF911 or #FukYoFlag and #BlackLivesMatter movements called for the lynching and hanging of white people and cops. They encouraged others on a radio show Tuesday night to turn the tide and kill white people and cops to send a message about the killing of black people in America.One of the F***YoFlag organizers is called Sunshine. She has a radio blog show hosted from Texas called, Sunshine s F***ing Opinion Radio Show. A snapshot of her #FYF911 @LOLatWhiteFear Twitter page at 9:53 p.m. shows that she was urging supporters to Call now!!
……

AFTER

law enforcement alert threats cops whites by black lives matter terrorists
comment expected barack obama fukyoflag black lives matter movements called lynching hanging white people cops encouraged radio tuesday night tide kill white people cops send message killing black people america fyoflag organizers called sunshine radio blog hosted texas called sunshine fing opinion radio snapshot twitter fiftythree aftermidday urging supporters…….

POS (Part of Speech) - Tagging

BEFORE

law enforcement alert threats cops whites by black lives matter terrorists
comment expected barack obama fukyoflag black lives matter movements called lynching hanging white people cops encouraged radio tuesday night tide kill white people cops send message killing black people america fyoflag organizers called sunshine radio blog hosted texas called sunshine fing opinion radio snapshot twitter fiftythree aftermidday urging supporters…….

AFTER

```
[('law', 'n'),
 ('enforcement', 'n'),
 ('alert', 'a'),
 ('threats', 'n'),
 ('cops', 'n'),
 ('whites', 'n'),
 ('blacklivesmatter', 'n'),
 ('terrorists', 'n'),
 ('video', 'n'),
 ('comment', 'n'),
 ('expected', 'v'),
 ('barack', 'n'),
 ('obama', 'n'),
,......
 ]
```

Lemmatization

BEFORE

law enforcement alert threats cops whites by black lives matter terrorists

comment expected barack obama fukyoflag black lives matter movements called lynching hanging white people cops encouraged radio tuesday night tide kill white people cops send message killing black people america fyoflag organizers called sunshine radio blog hosted texas called sunshine fing opinion radio snapshot twitter fiftythree aftermidday urging supporters…….

AFTER

law enforcement alert threat cop white black lives matter terrorist

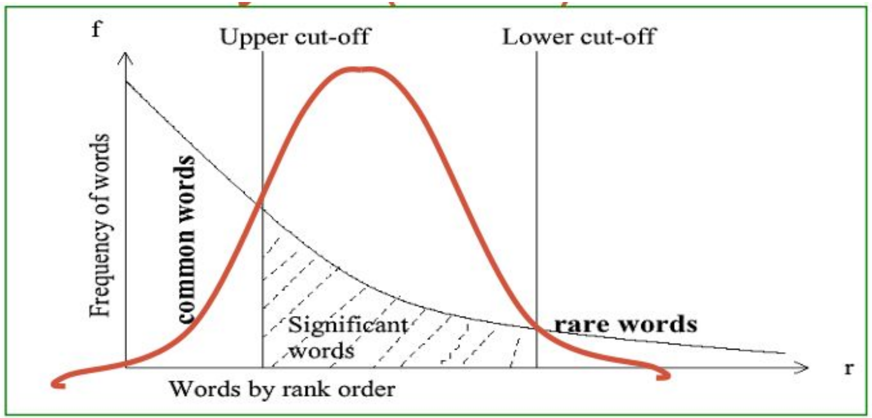
comment expect barack obama fukyoflag black lives matter movement call lynching hanging white people cop encourage radio tuesday night tide kill white people cop send message kill black people america fyoflag organizer call sunshine radio blog hosted texas call sunshine fing opinion radio snapshot fyf911 twitter fiftythree aftermidday urge supporter….

Stop Words Removal - Zipf's Law & Luhn's Analysis

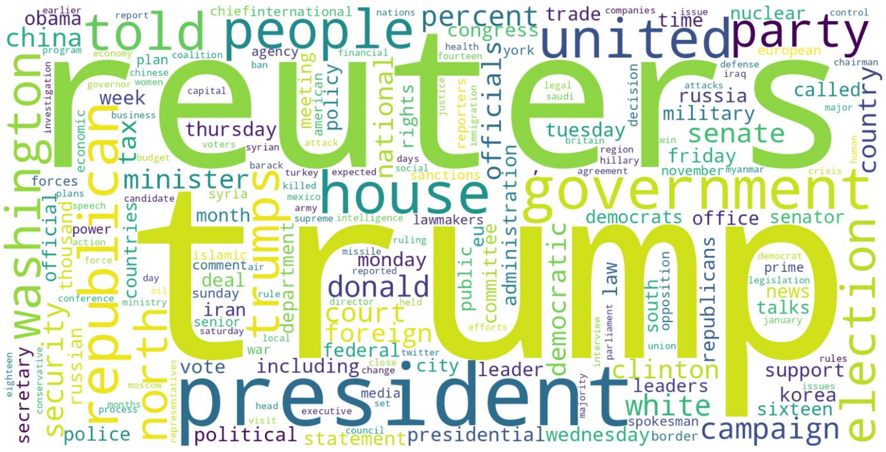
Why Stop Word Removal is used?
Source: QIAO ET AL., 2019 - Qiao, C. Xiong, Z. Liu, and Z. Liu, "Understanding the behaviors of bert in ranking," 2019.

What is Zipf's Law ?
Source: ZIPF, 1949 - G. K. Zipf, Human Behaviour and the Principle of Least Effort. Addison-Wesley, 1949.

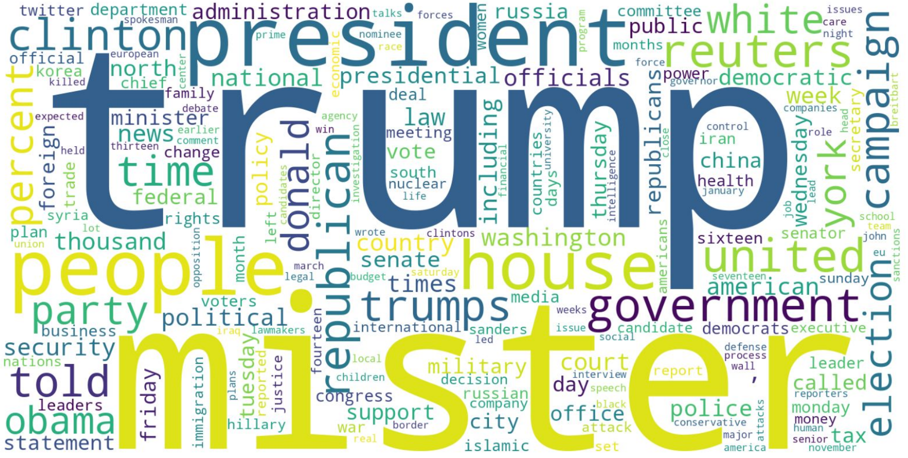
What is Luhn's analysis?
Source: LUHN, 1958 - H. P. Luhn, "The automatic creation of literature abstracts," IBM Journal of Research and Development, vol. 2, no. 2, pp. 159–165, 1958.



Most Frequent Words in Real News



WELFake Dataset



ISOT

Most Frequent Words in Fake News

WELFake Dataset

ISOT

2

Vectorization & Embeddings

Text Vectorization

Tokenize the text data using **Keras Text Vectorization**, converting words into sequences of integers. Then **padding or truncate** the sequences to a fixed length of 100.

PREPROCESSED NEWS

SIZE < 100

law enforcement alert threat cop white 911by blacklivesmatter fyf911 terrorist video comment expect barack obama fyf911 fukyoflag blacklivesmatter movement call lynching hanging white people cop encourage radio tuesday night tide kill white people cop send message kill black people america fyoflag organizer call sunshine radio blog hosted texas call sunshine fing opinion radio snapshot fyf911 twitter fifty-three aftermidday urge supporter....

VOCABULARY SIZE = 10K

SIZE = 100

[4721, 1733, 1004, 84, 401, 135, 278, 110, 591, 627, 4570, 682, 1119, 401, 176, 84, 67, 47, 41,
1200, 2619, 2062, 1441, 1391, 969, 333, 536, 3346, 28, 313, 230, 2, 2680, 307, 152, 110, 1200,
176, 27, 1391, 84, 521, 4299, 591, 35, 3155, 1882, 272, 95, 27, 2, 249, 262, 549, 400, 215, 16,
871, 493, 468, 1732, 168, 230, 272, 44, 47, 975, 77, 26, 44, 27, 1391, 1, 60, 274, 1678, 466,
3, 5, 319, 118, 48, 2585, 3053, 1163, 4150, 639, 4231, 537, 318, 2306, 0, 0, 0, 0, 0, 0, 0, 0, 0]

PADDING

Glove Embedding

GloVe (Global Vectors for Word Representation) embeddings are pre-trained word vectors that capture **semantic relationships** between words based on their co-occurrence statistics in large text corpora. Each word of a sentence is converted into vectors using context-based. Source: PENNINGTON ET AL., 2014 - GloVe: Global Vectors for Word Representation

SIZE = 100

[4721, 1733, 1004, 84, 401, 135, 278, 110, 591, 627, 4570, 682, 1119, 401, 176, 84, 67, 47, 41, 1200, 2619, 2062, 1441, 1391, 969, 333, 536, 3346, 28, 313, 230, 2, 2680, 307, 152, 110, 1200, 176, 27, 1391, 84, 521, 4299, 591, 35, 3155, 1882, 272, 95, 27, 2, 249, 262, 549, 400, 215, 16, 871, 493, 468, 1732, 168, 230, 272, 44, 47, 975, 77, 26, 44, 27, 1391, 1, 60, 274, 1678, 466, 3, 5, 319, 118, 48, 2585, 3053, 1163, 4150, 639, 4231, 537, 318, 2306, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Context Aware Embedding

M (Embedding Size) X N (Sentence Words)

[[-2.46450007e-01 4.52479988e-01 8.37199986e-01 1.30099997e-01 -1.07730001e-01 -2.36489996e-01 ....],
[3.51500005e-01 1.67349994e-01 -3.21289986e-01 5.49229980e-01 1.15829997e-01 5.87419987e-01 .....],
[3.63620013e-0 1.40389996e-02 3.54079992e-01 -9.76969972e-02 -9.76199985e-01 -5.58379984e+00 ....],
[-6.54269993e-01 1.99650005e-01 -2.74520010e-01 -4.54569995e-01 -6.27870020e-03 1.38549998e-01 ....],
[-2.67190002e-02 -6.72100019e-03 -5.70349991e-01 -3.69089991e-01 2.27440000e-01 -3.74390006e-01 ....],
[3.77070010e-01 2.15529993e-01 8.52970034e-02 1.57020003e-01 -7.48149991e-01 1.38549998e-01 ...],
......]

SIZE
100 X 100

4

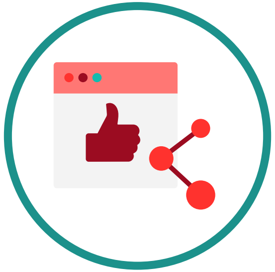
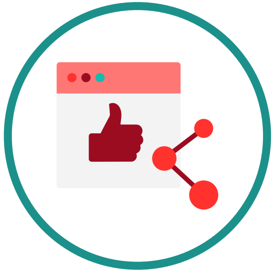
Classification Model

Workflow

Data Split	Train 80% Validation 20 % Test 20%	1
Embedding	Tokenization and Glove Embedding	2
Models	5 epochs with batch size 64 for each model	3

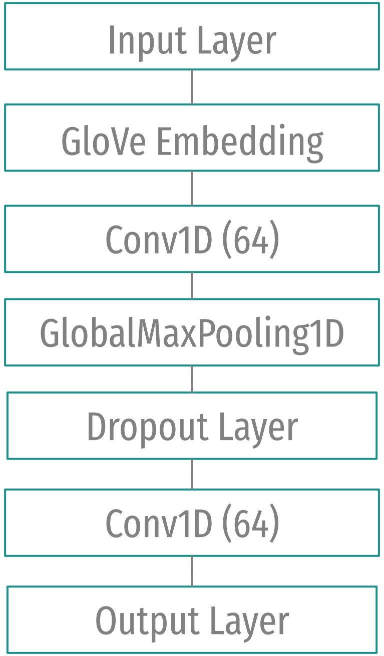
Classification Models

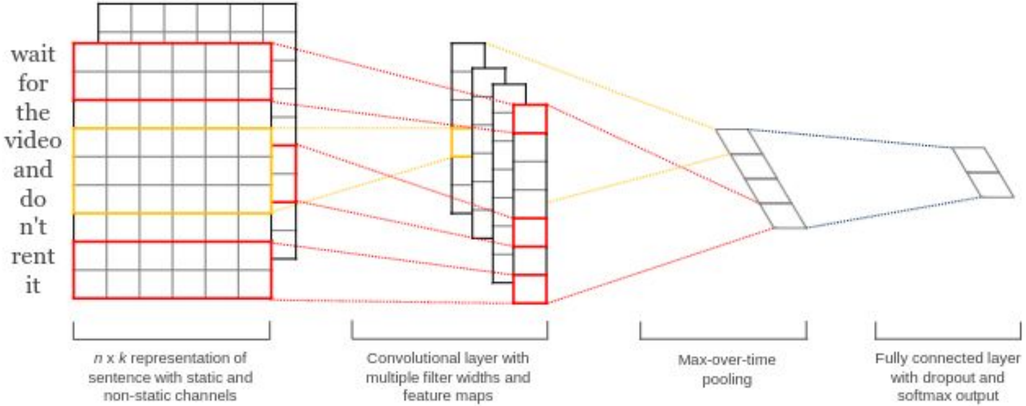
Each model will be trained using
EPOCHS = 5 BATCH SIZE = 64

1	2	3	4
			
CNN	**LSTM**	**BI-LSTM**	**BERT**
Simple Keras Neural Network	Optimized RNN which consider the sequentiality of words	Optimized RNN which consider the sequentiality of words in both direction	Use of a BERT as classifier in Fine-Tuning way

1 CNN

A CNN with 1-D convolution is a one-dimensional convolutional layer used for processing temporal data, **capturing local pattern and features**. It is commonly used in sequence processing tasks such as natural language processing (NLP) as done in the paper of [Kine 2014 -Convolutional Neural Networks for Sentence Classification](https://arxiv.org/abs/1408.5882)





Source: https://arxiv.org/abs/1408.5882

2

LSTM

LSTM networks Hochreiter et al. - 1997 , an optimization of RNN, are highly effective for processing sequence data, including text. They excel at retaining information over long sequences, enabling them to **understand sentence context** and assist in text classification, capturing **long-term dependencies** and **sequential information**.

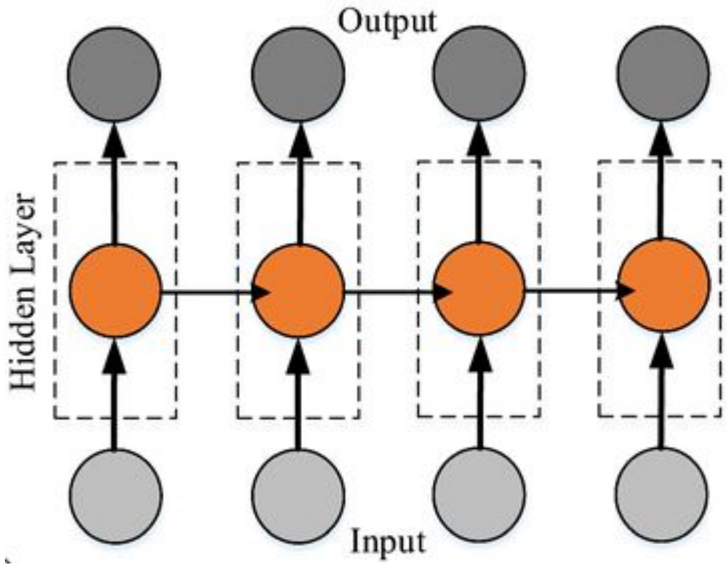
Input Layer

GloVe Embedding

LSTM

Dropout Layer

Output Layer



Source: LSTM - Architecture - Hochreiter et al.- 1997

3

BI-LSTM

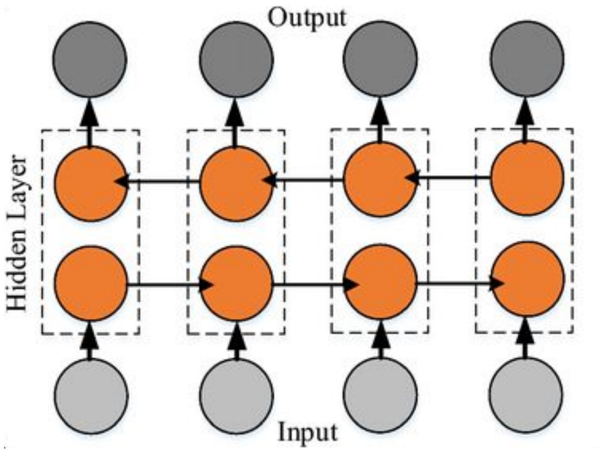
The BI-LSTM (Bidirectional LSTM) architecture, introduced by Graves et al. in 2013, employs Bidirectional Layers to comprehend sentences in both forward and backward directions. This approach enhances text classification by **capturing diverse patterns from both ends of the sequence**, leading to a **deeper understanding of the context**.

Input Layer

GloVe Embedding

Bidirectional - LSTM

Output Layer



Source: BI-LSTM - Architecture - Graves et al. in 2013

4

BERT

BERT (Devilin et a. In 2019) offers **state-of-the-art performance in various classification tasks**, including those cited in *Rogers et al. in 2019*. Utilizing fine-tuning, the small bert-uncased variant provides a lightweight and uncased version of BERT.

With BERT, **bidirectional context** is leveraged through **deep contextualized representations**. It employs a specialized tokenizer called Wordpiece, developed by Google, which breaks words into smaller units known as 'word pieces.' Schuster Et Al. in 2012

When using BERT, **partially pre-processed data is preferred over lemmatization**. Studies like Kutuzov et al. in 2019 , indicate that lemmatization does not yield significant improvements in simple morphological language like English.

Input Layer

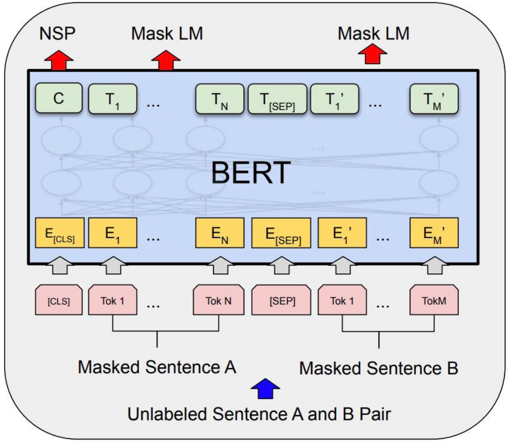
Preprocessing Layer

Encoder

Dropout Layer

Output Layer

→ BERT-EN-UNCASED



Source: Devilin et a. In 2019

Model Comparison

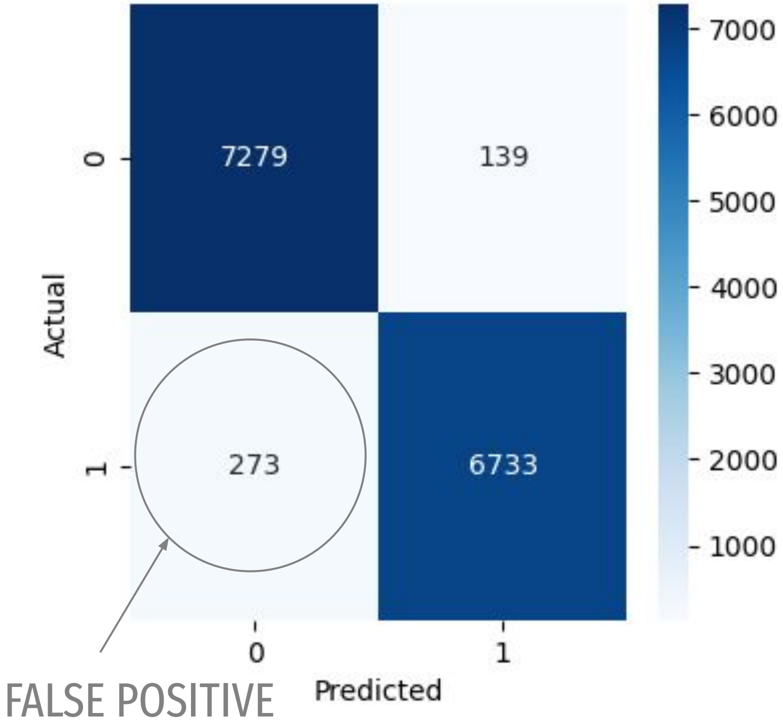
Why CNN are better than LSTM in classification tasks?
AJIK ET AL., 2023 - Fake news detection using optimized cnn and lstm techniques

		Accuracy	Precision	Recall	F1-Score	Training Time	Number Parameters
WELFAKE	GLOVE + CNN	0.91	0.92	0.91	0.91	24.5s	1.036.489
	GLOVE +LSTM	0.75	0.81	0.75	0.75	18m 32s	1.017.257
	GLOVE +BI-LSTM	0.94	0.94	0.94	0.94	35m 44s	1.482.201
	BERT	0.97	0.97	0.90	0.97	36m 28s	28.764.162
ISOT	GLOVE +CNN	1.00	1.00	1.00	1.00	13.7s	1.036.489
	GLOVE + LSTM	0.95	0.95	0.95	0.95	12m 27s	1.017.257
	GLOVE +BI-LSTM	1.00	1.00	1.00	1.00	23m 26s	1.482.201
	BERT	1.00	1.00	1.00	1.00	21m 50s	28.764.162

BERT - WELFake

BERT Fine-Tuned appears to outperform other models, leveraging the **full context** of a text.

	Precision	Recall	F1-Score	Support
FAKE	0.96	0.98	0.97	7418
REAL	0.98	0.96	0.97	7006
ACCURACY			0.97	14424
MACRO AVG	0.97	0.97	0.97	14424
WEIGHTED AVG	0.97	0.97	0.97	14424



Is possible to improve the generalization of the model?

Methodological Approach

How evaluate the models?

1

Basic Models

Evaluate model on both dataset (ISOT and WELFake)

2

Data Augmentation

Evaluate the models adding noise through Back Translation

3

Dataset Combination

Evaluate the models combining different sources, increasing training size

Does each Classifier perform well increasing the data variability ?

5

Data Augmentation & Combination

Data Augmentation - Back Translation

Back translation is a data augmentation technique commonly used in natural language processing (NLP). It involves translating a piece of text from its original language to another language and then translating it back to the original language. This process introduces **variations in the text while retaining its original meaning.** Back Translation help improve the **robustness and generalization** of NLP models (Shorten er al. 2021-Journal of Big Data)

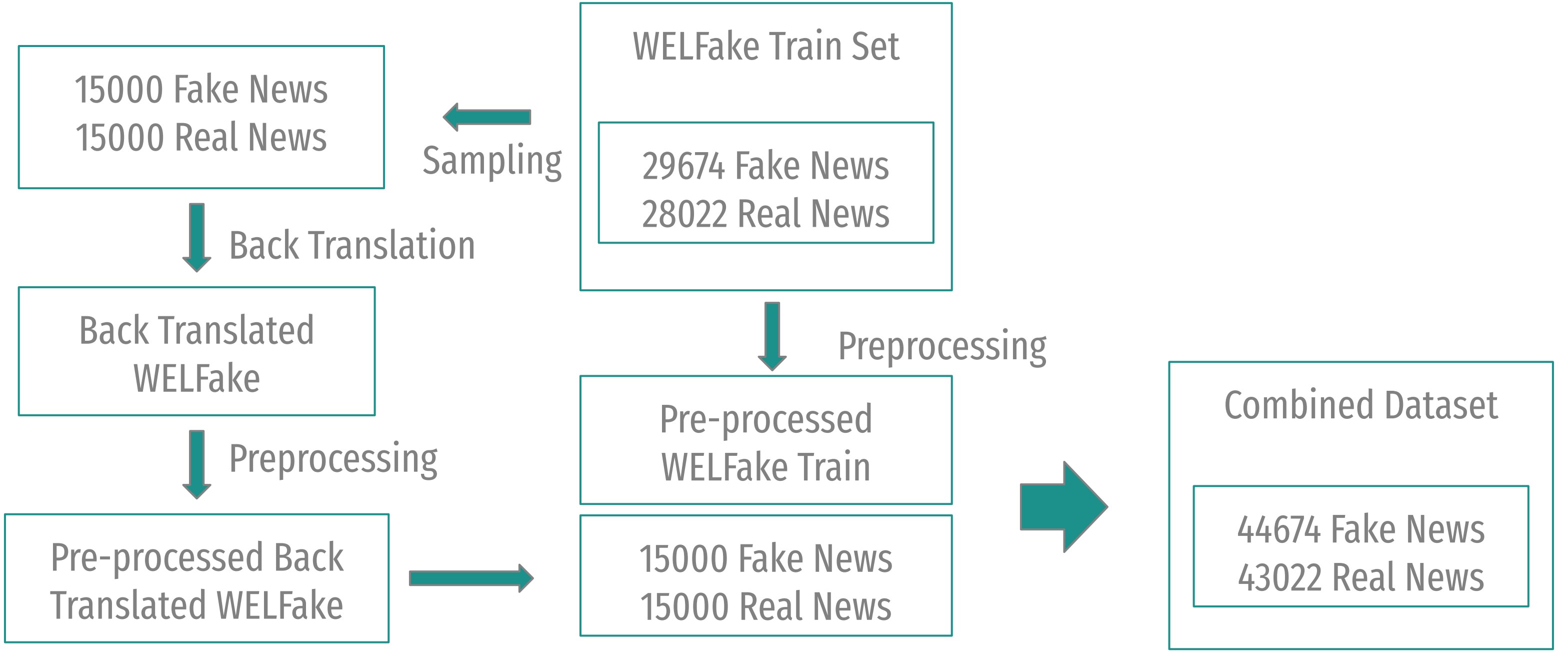
Back Translation is useful even in the case of Neural Embedding Corbeil et al. 2020

Considering that the text in the dataset are in English, using **French as the intermediate language** in back translation can indeed yield good results in data augmentation for natural language processing tasks

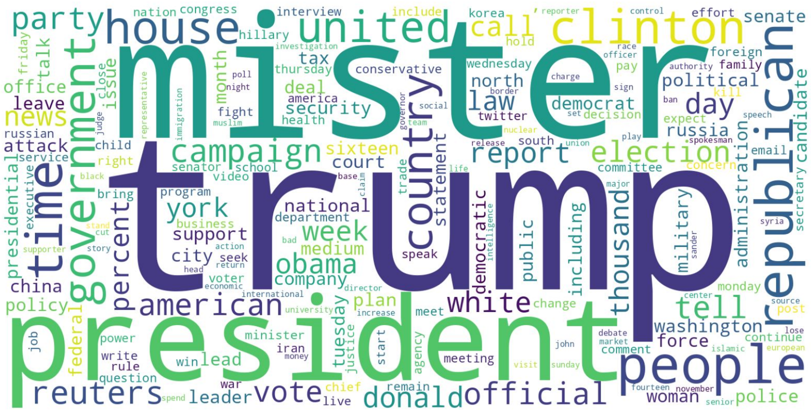
WELFake Dataset	French	English
White House blames Iran-backed militia for deadly drone strike The White House has blamed an Iran-backed militia for a deadly drone strike on an American base...	**La Maison Blanche accuse une milice soutenue par l'Iran d'être à l'origine d'une attaque de drone** meurtrière La Maison Blanche a accusé une milice soutenue par l'Iran d'être à l'origine d'une attaque de drone meurtrière sur une base...	**The White House accuses an Iranian-backed militia of being behind a deadly drone attack** The White House has accused an Iranian-backed militia of being behind a deadly drone attack on a U.S. military base in...

Data Augmentation - Back Translation

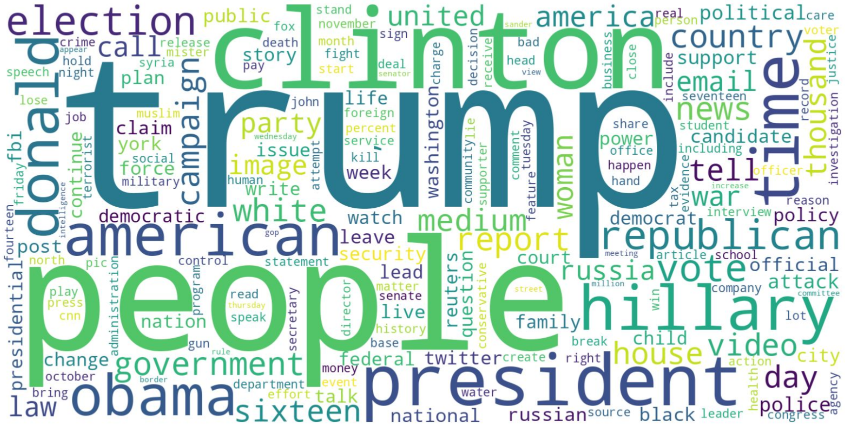
The basic preprocessed WELFAKE train-set will be combined with half of the back-translated dataset. This combination will enhance the **variability of the data**, allowing us to test the model's ability to **recognize various patterns within the data**.



Combined Train Set - Most Frequent Words



Real News



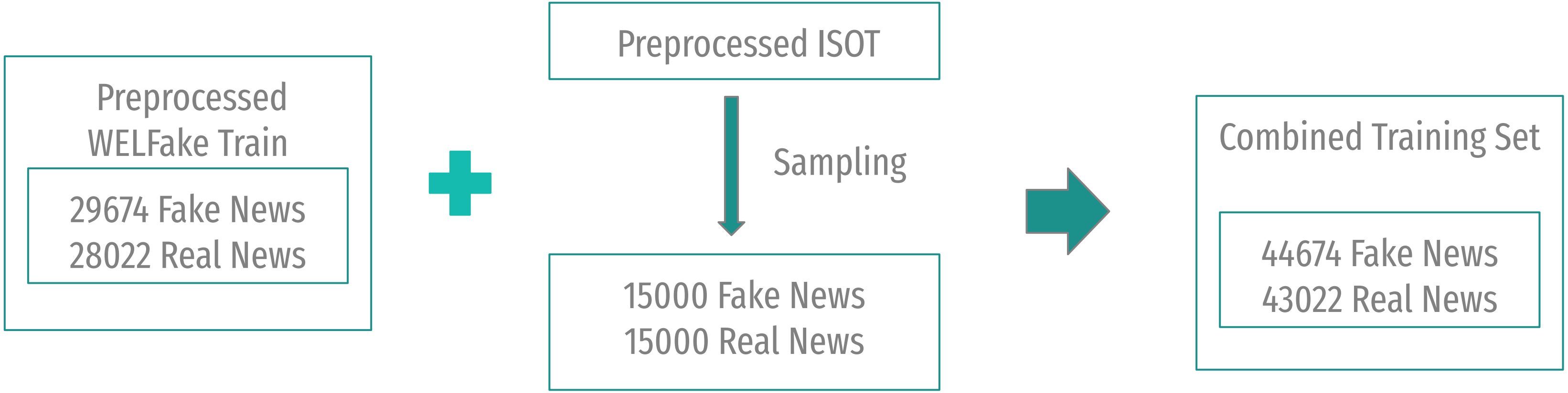
Fake News

Increase the variability of the words in the training set

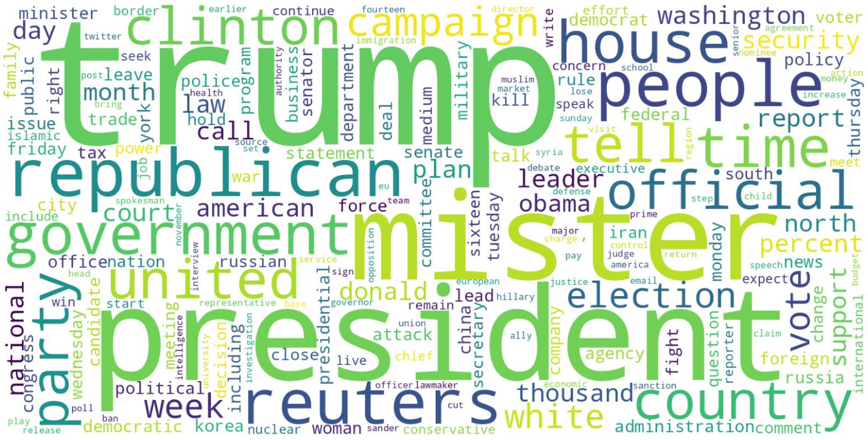
Dataset Combination WELFake + ISOT

Combining datasets serve as a form of data augmentation, by combining compatible datasets from different sources, you can increase the diversity and richness of your training data, which can help **improve the robustness and generalization** of models

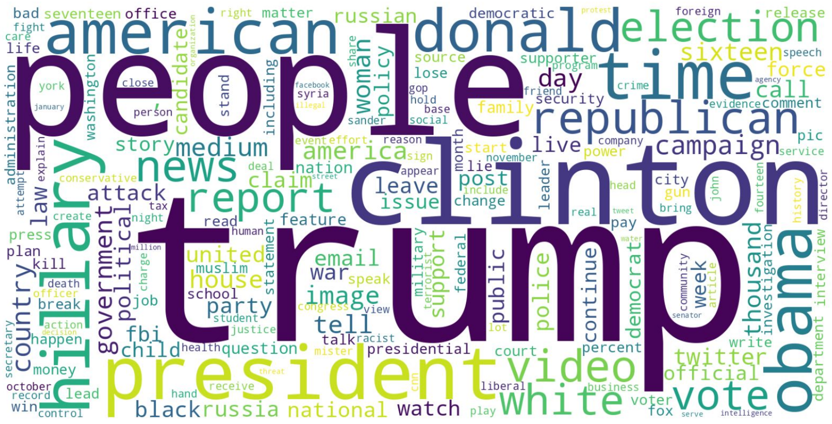
Combination of the WELFake and ISOT datasets aim to **increase data variability** and assess the model's ability to work with data from **different sources**, increasing the training size.



Combined Train Set - Most Frequent Words



Real News



Fake News

Increase the variability of the words in the training set

Back Translation vs Dataset Combination

		Accuracy	Precision	Recall	F1-Score	Training Time	Parameters Number
BACK TRANSLATION	GLOVE + CNN	0.83	0.83	0.83	0.83	3m 59s	1.036.489
	GLOVE + LSTM	0.73	0.74	0.73	0.73	27m 25s	1.017.257
	GLOVE + BI-LSTM	0.84	0.84	0.84	0.84	53m 53s	1.482.201
	BERT	0.92	0.92	0.92	0.92	52m 3s	28.764.162
DATASET COMBINATION	GLOVE + CNN	0.92	0.92	0.92	0.92	42s	1.036.489
	GLOVE + LSTM	0.81	0.82	0.80	0.80	29m 26s	1.017.257
	GLOVE + BI-LSTM	0.94	0.94	0.94	0.94	54m	1.482.201
	BERT	0.97	0.97	0.97	0.97	53m 3s	28.764.162

6

Knowledge Distillation

Knowledge Distillation

In this last step, a custom neural network is created, designed to have the **minimum number of parameters** while trying **not to excessively impact the accuracy.** The created model is as follows: In an attempt to achieve the performance of more complex models, the technique of "knowledge distillation" is also employed. It is useful for transferring knowledge from a large and complex model (teacher model) to a smaller and simpler model (student model). This techniques is effective in NLP task, as presented in Sun et al. del 2019

TEACHER MODEL
FINE TUNED BERT

Preprocessed
Train Set

LARGE COMPLEX MODEL
28.764.162 parameters

Knowledge Distillation

COMPRESSED MODEL

STUDENT MODEL
SIMPLE CNN 1D

Preprocessed
Train Set

SMALL SIMPLE MODEL
7.361 parameters

Knowledge Distillation - WELFAKE & ISOT

		Accuracy	Precision	Recall	F1-Score	Training Time	Number Parameters
WELFAKE	BERT FINE TUNED	0.97	0.97	0.97	0.91	36m 28s	28.764.162
	STUDENT NETWORK	0.90	0.91	0.90	0.90	41.9s	7.361
	DISTILLED STUDENT NETWORK	0.90	0.91	0.90	0.90	18m 10s	7.361
ISOT	BERT FINE TUNED	1.00	1.00	1.00	1.00	21m 50s	28.764.162
	STUDENT NETWORK	1.00	1.00	1.00	1.00	21.4s	7.361
	DISTILLED STUDENT NETWORK	1.00	1.00	1.00	1.00	11m 24s	7.361

7

Analysis of Results

Final Analysis

For each of the previous analysis, the best model appears to be **BERT**. With its ability to recognize the entire context of the sentence, it performs better during classification.

1. **WELFake**
   Baseline Dataset
   96%

2. **ISOT**
   Second considered Dataset
   100%

3. **Dataset Combination**
   Using different sources
   97%

4. **WELFake Back Translation**
   Increase train set variability
   92%

Conclusion

The previous analysis enables us to compare and evaluate the performance of different Deep Learning and NLP techniques on various datasets. It helps determine which technique is the most effective across different datasets for fake news classification, evaluating the model's ability to generalize to diverse data using data augmentation.

Analysis Fake News Language

Fake News use a specific language?

Preprocessing Impact

Does NLP pipeline improve the performance?

Compare different Classifier

Which model achieves the best performance?"

Augmentation Impact

Augmentation Techniques Improve the performance?

Conclusion

Evaluate Different Model
Deep Learning and NLP techniques

Importance of Preprocessing
Quality of the input data

Data augmentation
Increase the variability of the model

Model Distillation
Ability to transfer knowledge

Feature Development
Use larger model and analyze fake news length

Thank you for your attention

Fake News Detection
Advanced Machine Learning
Academic Year 2023-2024

Nicolò Urbani 856213
Mattia Piazzalunga 851931

Bibliography

[1] Oxford English Dictionary. (2023, September) fake news (n.). [Online]. Available: https://doi.org/10.1093/OED/3351660493
[2] Eurostat, “How many people verified online information in 2021?” December 2021. [Online]. Available: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20211216-3
[3] P. K. Verma, P. Agrawal, I. Amorim, and R. Prodan, “Welfake: Word embedding over linguistic features for fake news detection,” IEEE Trans-actions on Computational Social Systems, vol. 8, no. 4, pp. 881–893, 2021.
[4] H. Ahmed, I. Traore, and S. Saad, “Detection of online fake news using n-gram analysis and machine learning techniques,” in Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, I. Traore, I. Woungang, and A. Awad, Eds. Cham: Springer Interna- tional Publishing, 2017, pp. 127–138.
[5] Y. Qiao, C. Xiong, Z. Liu, and Z. Liu, “Understanding the behaviors of bert in ranking,” 2019.
[6] G. K. Zipf, Human Behaviour and the Principle of Least Effort. Addison-Wesley, 1949.
[7] H. P. Luhn, “The automatic creation of literature abstracts,” IBM Jour-nal of Research and Development, vol. 2, no. 2, pp. 159–165, 1958.
[8] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://aclanthology.org/D14-1162
[9] Y. Kim, “Convolutional neural networks for sentence classification,” CoRR, vol. abs/1408.5882, 2014. [Online]. Available: http://arxiv.org/abs/1408.5882
[10] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, p. 1735–1780, Nov 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

Bibliography

[11] T. Mikolov, M. Karafiát, L. Burget, J. H. ernocký, and S. Khudanpur, “Recurrent neural network based language model,” in Interspeech, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID: 17048224
[12] A. Graves, A. Mohamed, and G. E. Hinton, “Speech recognition with deep recurrent neural networks,” CoRR, vol. abs/1303.5778, 2013. [Online]. Available: http://arxiv.org/abs/1303.5778
[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in North American Chapter of the Association for Computational Linguistics, 152019. [Online]. Available: https://api.semanticscholar.org/CorpusID: 52967399
[14] A. Rogers, O. Kovaleva, and A. Rumshisky, “A primer in bertology: What we know about how BERT works,” CoRR, vol. abs/2002.12327, 2020. [Online]. Available: https://arxiv.org/abs/2002.12327
[15] M. Schuster and K. Nakajima, “Japanese and korean voice search,” in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 5149–5152.
[16] A. Kutuzov and E. Kuzmenko, “To lemmatize or not to lemmatize: How word normalisation affects elmo performance in word sense disambiguation,” ArXiv, vol. abs/1909.03135, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202540040
[17] C. Shorten, T. M. Khoshgoftaar, and B. Furht, “Text data augmentation for deep learning,” Journal of Big Data, vol. 8, no. 1, p. 101, 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00492-0
[18] J.-P. Corbeil and H. A. Ghadivel, “Bet: A back translation approach for easy data augmentation in transformer-based paraphrase identification context,” 2020.
[19] H. Sun, X. Tan, J.-W. Gan, S. Zhao, D. Han, H. Liu, T. Qin, and T.-Y. Liu, “Knowledge distillation from bert in pre-training and fine- tuning for polyphone disambiguation,” in 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019, pp. 168–175.
[20] E. Ajik, G. Obunadike, and F. Echobu, “Fake news detection using optimized cnn and lstm techniques,” Journal of Information Systems and Informatics, vol. 5, no. 3, pp. 1044–1057, Aug. 2023. [Online]. Available: https://www.journal-isi.org/index.php/isi/article/view/548