Fake News Classification

UNIVERSITÄT BONN

MA-INF 4232 - Lab Information Retrieval in Practice

Summer Semester 2021

Supervisor: Dr. Elena Demidova, Dr. Ran Yu

Presenter: Rohil Rao (3299480)

Motivation

- **Fake News**: Clickbait, Satire, Propaganda, Hoax, Sloppy journalism
- **Threats:** Misinformation, Mistrust in public institutions
- Social media, Big Data
- Expert fact-checkers
- Automation can help professionals/users

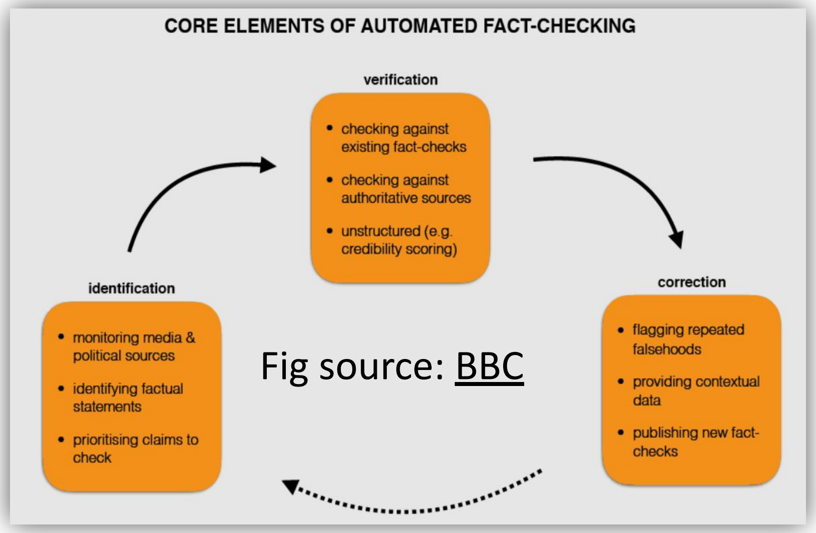


Figure from Graves 2018 [1]

Problem Description

- Task: Binary Classification of News (Real or Fake) for English Text

- Research Questions:

  1. How well can automated methods **perform** on the given task?
  2. **Comparison** of ML and Deep Learning model performance.
  3. What **feature engineering** methods can be used to extract insights?
  4. How can we **explain** our model predictions?

Progress Overview

APRIL-MAY:
Prepare Datasets, Literature review, Word Embeddings, Baseline models

JUNE:
Re-iterations, Model Tuning, Deep Learning, Explainability, Feature Engineering, Web Applications

JULY:
Final touches, Report

Outline

- Dataset preparation
- ML & DL models
- RQ3: **feature engineering** methods
- RQ4: **explaining** model predictions
- Conclusions

Datasets

	Datasets	Source	Samples	Note
1	Getting Real about FN	Kaggle - Risdal	13K	Highly skewed, Scraped using BS Detector tool
2	FN Detection (DS-1)	Kaggle - Jruvika	3K	Stratified subset of above. Used in [2]
3	FN UTK (DS-2)	Kaggle - UTK	20K	Kaggle competition dataset. Used in [3]
4	ISOT FN (DS-3)	University of Victoria	44K	Reuters, Politifact. Used in [3]
5	FakeNewsNet	Kai Shu github	18K	Tweets, Social, Spatiotemporal features. [4]

- Repository of 35 fake news datasets:
  - **Short claims**: FEVER, LIAR
  - **Social media texts**: FacebookHoax, BuzzFace
  - **Other languages**: Spanish, Arabic

Preparation

1. Lowercasing
2. Remove missing, duplicates
3. Remove outlier articles, other languages
4. Fix contractions (Don’t – do not)
5. Removing special characters
6. Stopword-removal
7. Tokenization and Lemmatization
8. Stratify

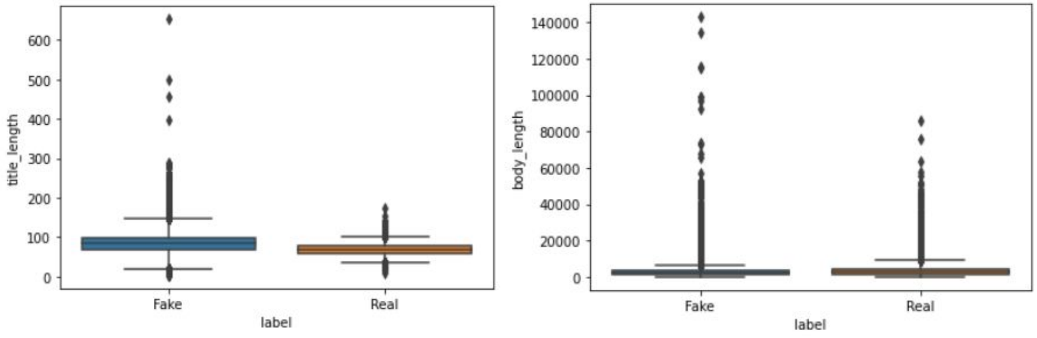


Fig: Boxplot lengths of title and body

DS-4	Before	After
Fake	36,031	27,907
Real	33,676	27,907
Total	69,707	55,814

Table: Dataset-4 before and after pre-processing

Word Representation

Term Freq - Inverse Document Freq : Accounts for term frequency in documents and also gives higher weight to rare terms.

Key points: 70:30 train-test split, 5 fold cross validation

Results for ML models on Dataset 4 (DS-4):

Model	Acc	AvgP	AvgR	AvgF1
Logistic Regression	0.91	0.91	0.91	0.91
Naive Bayes	0.85	0.85	0.85	0.85
Decision Trees	0.86	0.86	0.86	0.86
Random Forest	0.88	0.88	0.88	0.88
Adaboost	0.88	0.88	0.88	0.88

Word Representation

Word2Vec: Calculates probability of word based on neighbouring words for entire corpus.

Architecture and training:

- 3 layers (Embedding, LSTM, Dense output)
- Optimizer: Adam
- Loss: Binary Cross Entropy
- Train-test split: 70-30
- Epochs: 5

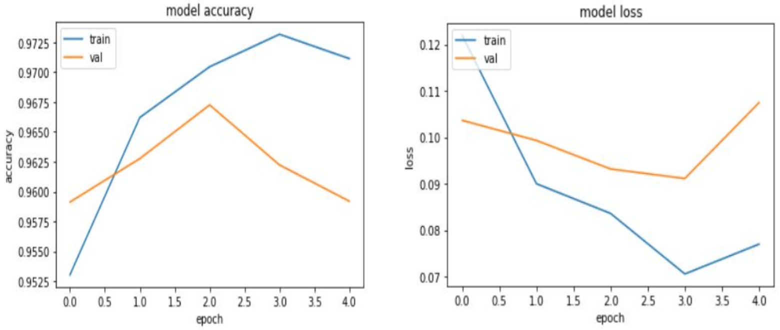
Result:

Model	Acc	AvgP	AvgR	AvgF1
W2V + LSTM	0.94	0.94	0.94	0.94

```
print("LENGTH",len(w2v_model.wv.__getitem__("facebook")))
w2v_model.wv.most_similar('facebook', topn=10)

LENGTH 100
[('reddit', 0.6705176830291748),
 ('instagram', 0.6616578102111816),
 ('google', 0.6589418649673462),
 ('snapchat', 0.6459718942642212),
 ('4chan', 0.6427757740020752),
```

Fig: similar words in corpus for trained model



Figs: Accuracy & Loss for 5 epochs

Word Embedding

Word2Vec: Calculates probability of word based on neighbouring words for entire corpus.

STEPS:

1. Convert text to lists of sentences (list of words)
2. Selected embedding dimension (100)
3. Gensim CBOW model for building vocabulary
4. Texts to sequences
5. Select max-length of articles and padding
6. Obtain embedding weight matrix
7. Create model architecture and train

```
print("LENGTH",len(w2v_model.wv.__getitem__("facebook")))
w2v_model.wv.most_similar('facebook', topn=10)
```

```
LENGTH 100
[('reddit', 0.6705176830291748),
 ('instagram', 0.6616578102111816),
 ('google', 0.6589418649673462),
 ('snapchat', 0.6459718942642212),
 ('4chan', 0.6427757740020752),
```

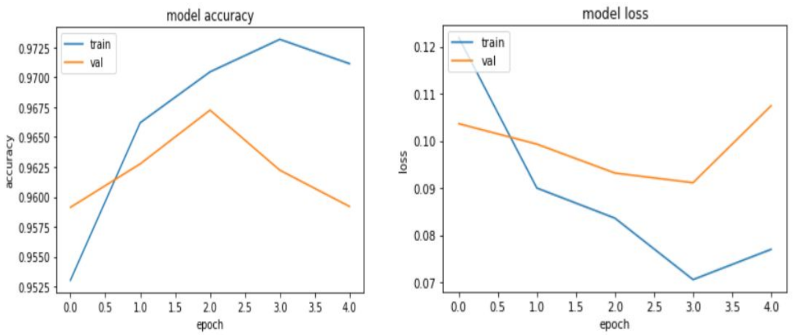
Fig: similar words in corpus for trained model

Model and Results

Architecture details:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 1000, 100)         22531200
_________________________________________________________________
lstm (LSTM)                  (None, 128)               117248
_________________________________________________________________
dense (Dense)                (None, 1)                 129
=================================================================
Total params: 22,648,577
Trainable params: 117,377
Non-trainable params: 22,531,200
```

Optimizer: Adam
Loss: Binary Cross Entropy
Train-test split: 70-30
Epochs: 5



Figs: Accuracy & Loss for 5 epochs

	precision	recall	f1-score	support
0.0	0.92	0.96	0.94	8298
1.0	0.96	0.92	0.94	8500
accuracy			0.94	16798
macro avg	0.94	0.94	0.94	16798
weighted avg	0.94	0.94	0.94	16798

Fig: Classification report for test set

RQ3: Feature Engineering

- **S1: Horne and Adali (2017)**: This Just In: Fake News Packs a Lot in Title
  - Smaller dataset (4.5K samples)
  - Fake news has Simpler, Repetitive Content in Body
  - Argue that title is more important for classification

- **S2: Shrestha & Spezzano (2021):** Reproducibility Study
  - Larger dataset FakeNewsNet (18K samples)
  - Compared more models
  - Confirm title features better but not in all cases

- **Our work**:
  - Larger dataset (60K samples), open-source tools instead of Linguistic Inquiry Word Count (LIWC).

Features

1. **STYLISTIC:** Text syntax, style, grammar
   - Word and Sentence count, Words-per-sentence,
   - Parts-of-speech counts (Nouns, Personal pronouns etc. to obtain 38 features) using NLTK

2. **PSYCHOLOGICAL:** Capture overall sentiment and emotions
   - VADER for overall sentiment, TextBlob for Subjectivity and Polarity
   - Empath: 200 pre-validated topics or emotions, and has high correlation with LIWC features

3. **COMPLEXITY:** Readability of text
   - Flesh Kincaid Grade Level (FK), Gunning Fog Index (GI), Simple Measure of Gobbledygook Index (SMOG) using TextStat
   - Type-Token Ratio for lexical diversity
   - Average word and sentence length

Statistical Results

- Two-sample T-tests (two-sided statistical tests) for the null hypothesis that two independent samples have identical averages at probability threshold less than 0.05

Textual characteristic	H & A (S1)	S & S (S2)	Ours	P-value (Ours)
Article length (Word Count), Avg. Word and Sentence length	R > F	R > F	R > F	2.05e-19
Article Lexical Diversity (TTR)	F > R	F > R	F > R	1.7e-45
Article Reqd. Reading education level	R > F	R > F	R > F	5.6e-19
Title length (Word count)	F > R	F > R	F > R	0.00e00
Title avg. word length	R > F	R > F	R > F	6.69e-04
Title Proper Nouns (NNP)	F > R	F > R	R > F	2.29e-06

Results

- Compare results for Title, Body as well as different features

Model	Stylistic		Psychology		Complexity		All Combined	
	AUROC	AvgP	AUROC	AvgP	AUROC	AvgP	AUROC	AvgP
Title (Adab)	**0.80**	**0.73**	0.81	0.72	**0.75**	**0.68**	0.85	0.70
Title (DT)	0.66	0.66	0.72	0.71	0.68	0.65	0.74	0.74
Title (RF)	0.80	0.73	**0.86**	**0.78**	0.73	0.67	**0.9**	**0.82**
Title (LR)	0.75	0.7	0.78	0.71	0.678	0.659	0.84	0.77
Body (Adab)	0.83	0.75	0.84	0.76	0.73	0.68	0.88	0.8
Body (DT)	0.69	0.69	0.69	0.69	0.62	0.62	0.73	0.73
Body (RF)	**0.86**	**0.78**	**0.88**	**0.8**	**0.76**	**0.69**	**0.91**	**0.83**
Body (LR)	0.81	0.73	0.85	0.77	0.67	0.63	0.89	0.81
Both (Adab)	0.88	0.81	0.87	0.79	0.81	0.75	0.91	0.84
Both (DT)	0.76	0.76	0.73	0.73	0.70	0.70	0.79	0.79
Both (RF)	**0.91**	**0.84**	**0.9**	**0.82**	**0.86**	**0.79**	**0.94**	**0.87**
Both (LR)	0.85	0.78	0.87	0.79	0.72	0.68	0.92	0.85

Fig: Our Results for feature engineering on combined dataset DS-4

	Baseline	Fake vs Real
Body	50%	71%
Title	50%	78%

Fig: Results from Horne and Adali (2017)

	PolitiFact		BuzzFeedNews		GossipCop	
Features	AUROC	AvgP	AUROC	AvgP	AUROC	AvgP
News body (SVM)	0.583	0.466	0.614	0.257	0.623	0.327
News body (LR)	0.855	0.809	0.728	0.351	0.703	0.437
News body (RF)	**0.911**	**0.878**	**0.785**	**0.417**	**0.782**	**0.630**
News Title (SVM)	0.833	0.804	0.669	0.317	0.588	0.309
News Title (LR)	0.849	0.813	0.787	0.423	0.663	0.380
News Title (RF)	**0.867**	**0.823**	**0.812**	**0.424**	**0.715**	**0.490**

Fig: Shrestha & Spezzano (2021)

RQ4: Explainability

LIME: Local Interpretable Model-Agnostic

- Can we trust model predictions?
- Intuition:
  - Treat all models black-box
  - Explains single instance
  - New dataset of perturbations using model
  - Fit simpler model to perturbed dataset

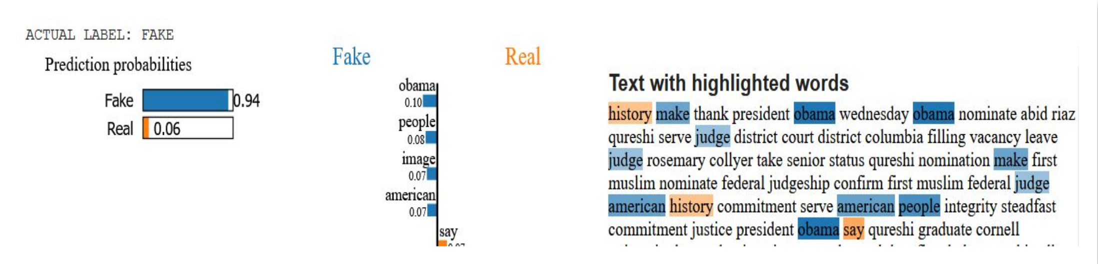


Fig: Example prediction for correctly predicted fake

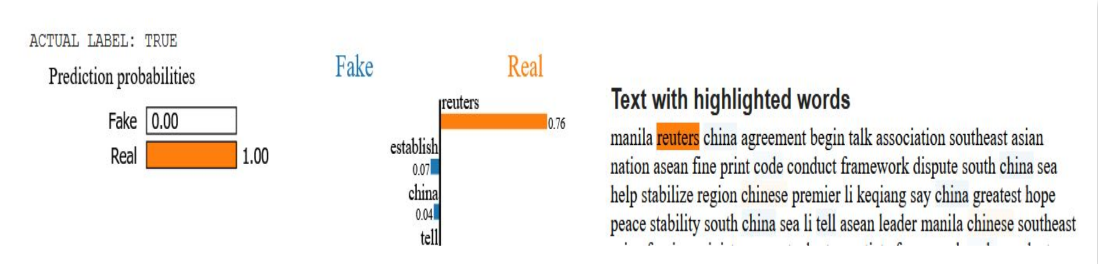
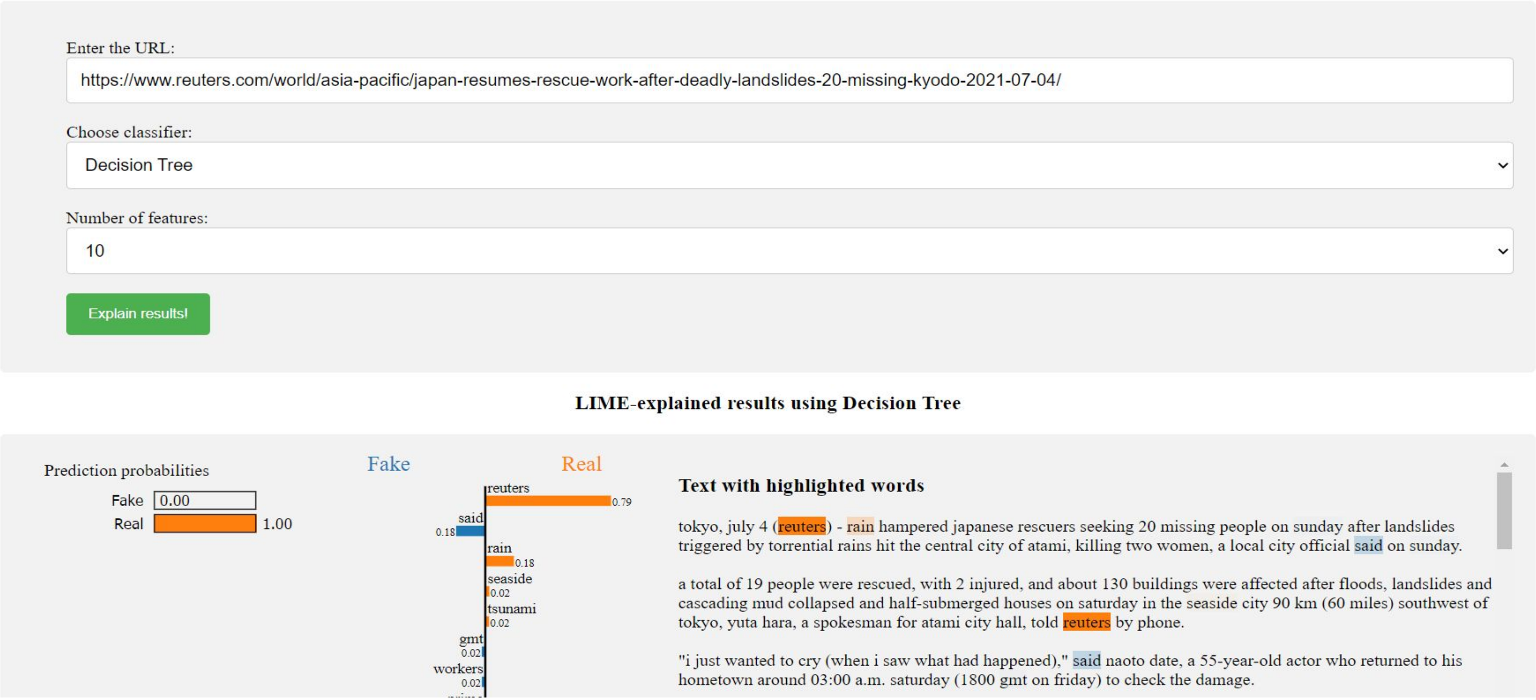


Fig: Example prediction for correctly predicted real

User Interface 1



Available online: https://irlab21-fakenews-explainer.herokuapp.com/

Summary

Conclusion:

- In our study title features were not found more important than body
- Explaining model predictions highly important

Future Scope:

- Datasets: spatiotemporal, image data
- Claim identification, relevant evidence retrieval
- Explainability methods: SHAP

References

[1] Lucas Graves. Understanding the Promise and Limits of Automated Fact-Checking. In Reuters Factsheet February 2018. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf

[2] Avinash Bharadwaj, Brinda Ashar, Parshva Barbhaya , Ruchi Bhatia , Zaheed Shaikh. Source Based Fake News Classification using Machine Learning. IJIRSET. 2020. http://www.ijirset.com/upload/2020/june/115_4_Source.PDF

[3] Ahmad, Iftikhar, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. "Fake News Detection Using Machine Learning Ensemble Methods." Complexity 2020. https://downloads.hindawi.com/journals/complexity/2020/8885861.pdf

[4] Kai Shu et al. (2018). Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. https://www.liebertpub.com/doi/full/10.1089/big.2020.0062

[5] Mikolov, Tomas; et al. (2013). "Efficient Estimation of Word Representations in Vector Space. https://arxiv.org/pdf/1301.3781.pdf%C3%AC%E2%80%94%20%C3%AC%E2%80%9E%C5%93

[6] Benjamin D. Horne, Sibel Adali (2017). This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. https://arxiv.org/abs/1703.09398

[7] Anu Shrestha, Francesca Spezzano (2021). Textual Characteristics of News Title and Body to Detect Fake News: A Reproducibility Study. https://www.springerprofessional.de/en/textual-characteristics-of-news-title-and-body-to-detect-fake-ne/19017670

[8] Ethan Fast et. al. Empath: Understanding Topic Signals in Large-Scale Text. https://hci.stanford.edu/publications/2016/ethan/empath-chi-2016.pdf

[9] Ribeiro et al. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. https://arxiv.org/abs/1602.04938