

M2 Software Project

OAKS

Fake News Detector

Anna MOSOLOVA
Elisa LUBRINI



UNIVERSITÉ
DE LORRAINE



September 28, 2021

Contents

1 Introduction

2 Literature Review

3 Our Solution

4 The Product

5 Next Steps

Table of Contents

1 Introduction

2 Literature Review

3 Our Solution

4 The Product

5 Next Steps

Introduction

News plays an important role in shaping people's beliefs and opinions. The spread of fake news is being increasingly problematic. [6]

Example

The past election cycle for the 45th President of the United States. [6]

NLP can help through **fake news detection**:

- **reminding** newsreaders to be alert
- providing information on the **accuracy level** of each case

Table of Contents

1 Introduction

2 Literature Review

3 Our Solution

4 The Product

5 Next Steps

Knowledge base approach

- DEAP-FAKED [4]
 - Dataset: Kaggle Fake News¹
 - biLSTM for article representation
 - ComplEx algorithm for knowledge base entity representation (WikiData)
 - Classification with Multilayer perceptron
 - F1: 88.66%

¹<https://www.kaggle.com/c/fake-news/overview>

Neural network approach

- Exploring Text-transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English [3]
 - Dataset: COVID19 Fake News Detection in English²
 - Bert, Ernie, Roberta, XL-net, and Electra ensemble with pseudo-labeling
 - F1: 98.59%

²<https://competitions.codalab.org/competitions/26655>

Linguistic features approach

- Linguistic feature based learning model for fake news detection and classification [1]
 - Dataset: Buzzfeed Political News Data, Random Political News Data
 - Features:
 - Syntax-based
 - Sentiment-based
 - Grammatical-based
 - Readability-based
 - Ensemble of Gaussian Naïve Bayes, Kernel Naïve Bayes, Linear SVM, Gaussian SVM
 - Accuracy: 72%

Table of Contents

1 Introduction

2 Literature Review

3 Our Solution

4 The Product

5 Next Steps

Our Solution

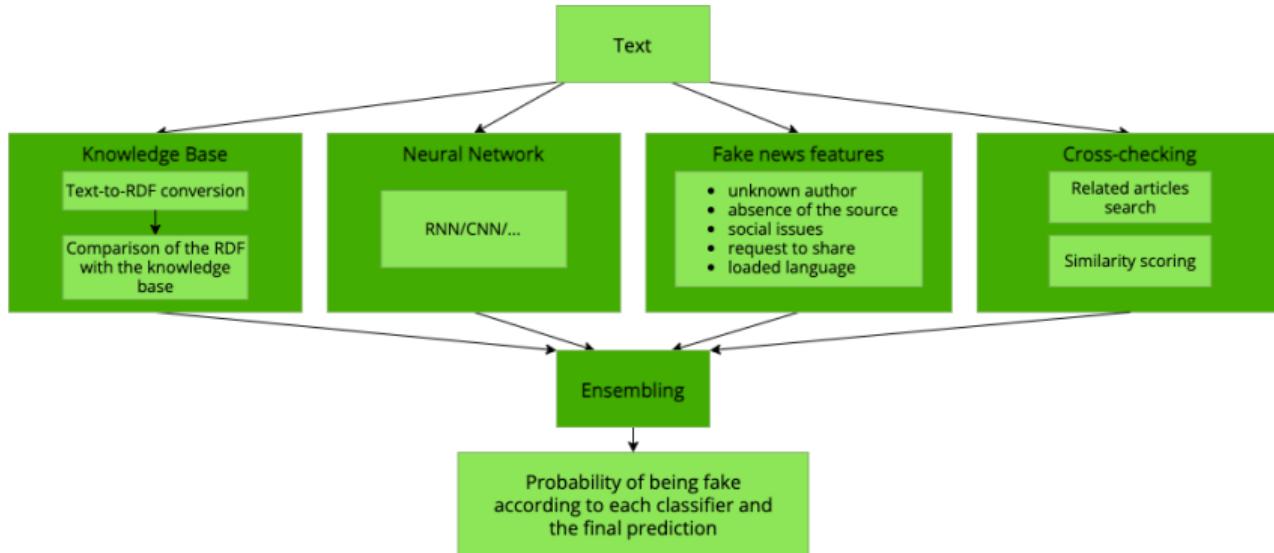


Figure: Ideal structure of the system's algorithm

Dataset

- LIAR dataset
- Training set: 10,269
- Validation set: 1,284
- Testing set: 1,283

Statement: *“Under the health care law, everybody will have lower rates, better quality care and better access.”*

Speaker: Nancy Pelosi

Context: on 'Meet the Press'

Label: False

Justification: Even the study that Pelosi's staff cited as the source of that statement suggested that some people would pay more for health insurance. Analysis at the state level found the same thing. The general understanding of the word “everybody” is every person. The predictions don't back that up. We rule this statement False.

Our Solution

a) Knowledge Base

- Text-to-RDF graph conversion [2]
- Comparison with the WikiData knowledge base

b) Neural Network

Modern neural network architectures that show SoTA results in the various NLP tasks

- GRU
- LSTM
- CNN
- Transfer learning

c) Feature-based Algorithm

As stated in [5], the following features are particularly common in fake news:

- unknown author
- absence of the source
- social issues
- requests to share the news
- loaded language usage

d) Cross-checking Algorithm

- search of the similar articles on GoogleNews
- verification if the information is not contradictory

Table of Contents

1 Introduction

2 Literature Review

3 Our Solution

4 The Product

5 Next Steps

The Product

Various options are possible:

- **Web App**
- Browser **Extension**
- **Smartphone App**

Front End

Germany: SPD intends to form coalition with Greens and liberals

Centre-left contender to replace Angela Merkel announces plan for 'social-ecological-liberal' alliance

Philip Oettermann in Berlin
@philipottermann
Mon 27 Sep 2021 15.51 BST

f t e

▲ Olaf Scholz sees three-way coalition after SPD's narrow German election win – video

The centre-left contender to fill Angela Merkel's shoes has announced his intention to forge a "social-ecological-liberal coalition" following Sunday's knife-edge German national vote, as momentum slips from the outgoing chancellor's own designated successor.

"The voters have made themselves very clear," Olaf Scholz of the Social Democratic party (SPD) said at a press conference on Monday morning. He pointed out that his centre-left party, the Greens and the pro-business Free Democratic party (FDP) had all picked up significant numbers of new votes at

Most viewed

- End to freedom of movement behind UK fuel crisis, says Merkel's likely successor
- Italy: bronze statue of scantily dressed woman sparks sexism row
- R Kelly found guilty on racketeering and sex trafficking charges
- Rolling Stones review - a funky, heavy first show without Charlie Watts
- 'Getting into Europe is a relief every time': an HGV driver reflects on UK crisis

Figure: Sample article

Front End

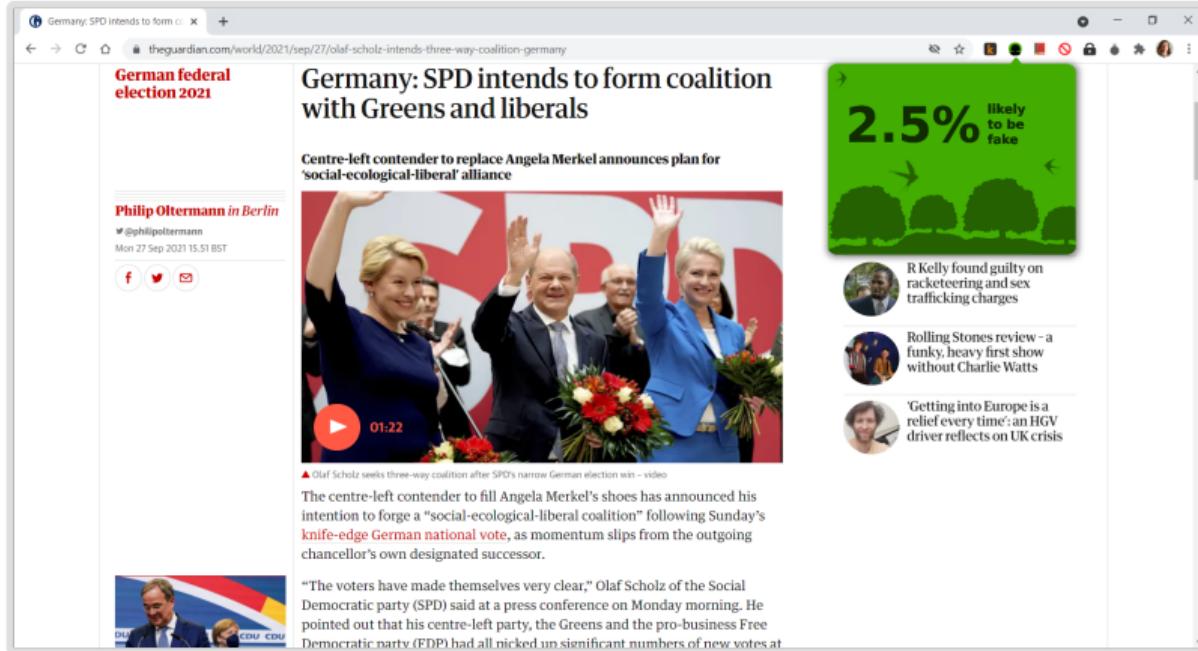


Figure: Draft mockup of the browser extension

Front End

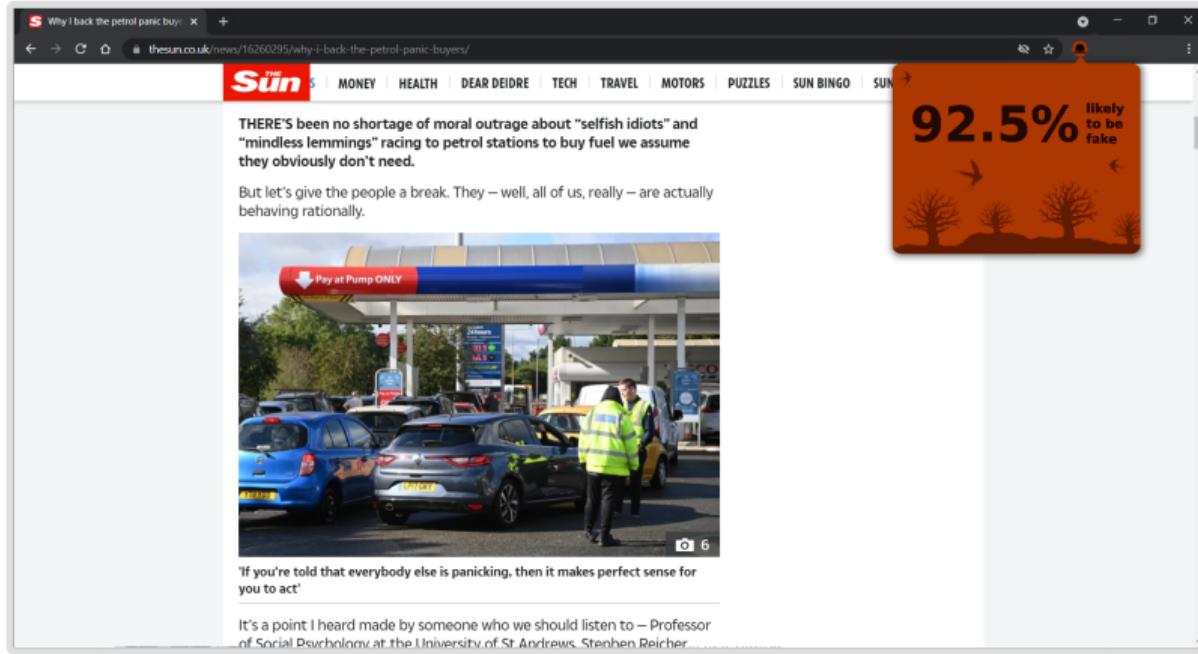


Figure: Draft mockup of the browser extension when detecting fake news

Back End

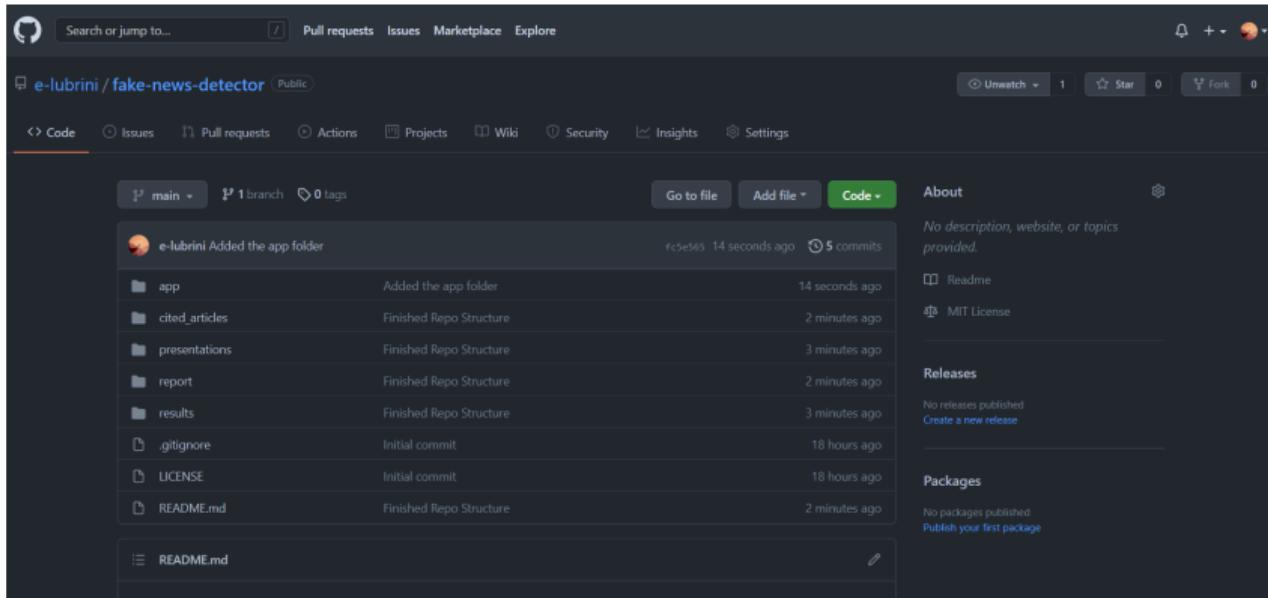


Figure: All our code will be available on our GitHub repo: **e-lubrini/fake-news-detector**

Table of Contents

1 Introduction

2 Literature Review

3 Our Solution

4 The Product

5 Next Steps

Coming Up...

For the next two weeks...

- Polish the **brand** (logo, ...)
- Research **ensembling** methods
- Create an approximate **timeline** for 4 app sections' workload
- Start **implementation**
- Find an **evaluation** method

Coming Up...

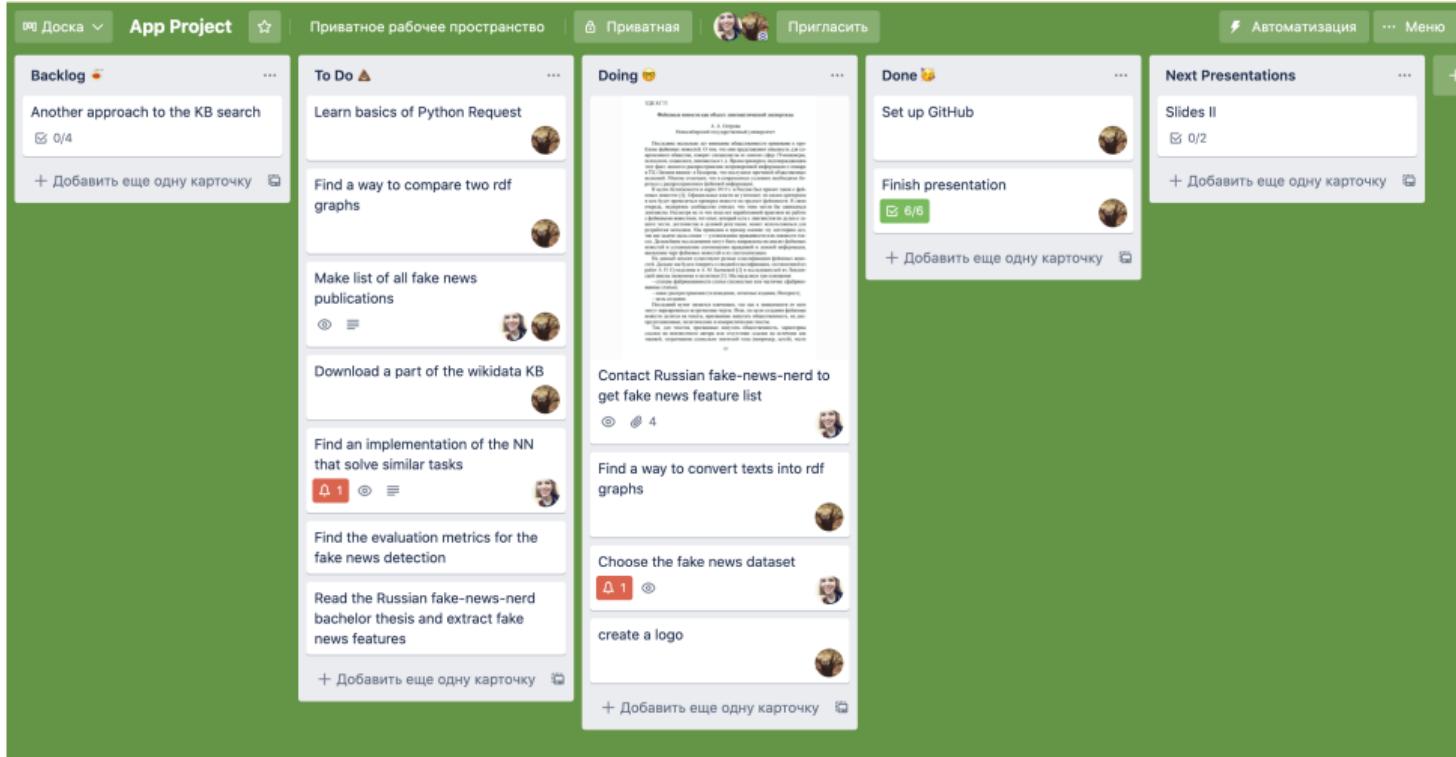


Figure: Task repartition between team members

Bibliography I

- [1] Anshika Choudhary and Anuja Arora. "Linguistic feature based learning model for fake news detection and classification". In: *Expert Systems with Applications* 169 (2021), p. 114171.
- [2] Francesco Draiaggio et al. "FRED: From natural language text to RDF and OWL in one click". In: vol. 7955. May 2013, pp. 263–267. ISBN: 978-3-642-38708-1. DOI: 10.1007/978-3-642-41242-4_36.
- [3] Xiangyang Li et al. "Exploring text-transformers in aaai 2021 shared task: Covid-19 fake news detection in english". In: *arXiv preprint arXiv:2101.02359* (2021).
- [4] Mohit Mayank, Shakshi Sharma, and Rajesh Sharma. "DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection". In: *arXiv preprint arXiv:2107.10648* (2021).
- [5] Anna Petrova. "Fake news as an object of the linguistic analysis". Russian. In: *58 International scientific student conference ISSC-2020: Applied linguistics*. Novosibirsk, 2020, pp. 19–20.

Bibliography II

- [6] Zhixuan Zhou et al. "Fake News Detection via NLP is Vulnerable to Adversarial Attacks". In: *Proceedings of the 11th International Conference on Agents and Artificial Intelligence* (2019). DOI: 10.5220/0007566307940800. URL: <http://dx.doi.org/10.5220/0007566307940800>.