

*M2 Software Project*

# Online App for Knowledge Substantiation

Anna MOSOLOVA  
Elisa LUBRINI



# Contents

1 Introduction

2 Related Work

3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

4 The App

5 Results

6 Conclusion

# Table of Contents

## 1 Introduction

## 2 Related Work

## 3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

## 4 The App

## 5 Results

## 6 Conclusion

## Definitions

- News piece → News content, in the form of an article, a video, etc.

# Introduction

## Definitions

- News piece → News content, in the form of an article, a video, etc.
- Fake news → news pieces that contain **false** or **misleading** information

# Introduction

## Definitions

- News piece → News content, in the form of an article, a video, etc.
- Fake news → news pieces that contain **false** or **misleading** information
- Fake news detection system → Binary **classifier**: is the news piece fake?

# Introduction



- ✓ Ensembling various modules  
→ 4 approaches

# Introduction



- ✓ Ensembling various modules  
→ 4 approaches
  
- ✓ Developing a user-friendly app  
→ Web App

# Introduction



- ✓ Ensembling various modules  
→ 4 approaches
- ✓ Developing a user-friendly app  
→ Web App
- ✓ Continuously improve/expand it  
→ plans from today

# Table of Contents

1 Introduction

2 Related Work

3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

4 The App

5 Results

6 Conclusion

# Related Work

A variety of approaches...

- Feature-based → *Syntax, Sentiment, Grammatical, and Readability [3]*

# Related Work

## A variety of approaches...

- Feature-based → *Syntax, Sentiment, Grammatical, and Readability* [3]
- Knowledge Bases
  - in Theory → ideal system: comparing article KB to reference KB [10]
  - in Practice → KG classification [6], DEAP-FAKED [9], TransE [2]

# Related Work

## A variety of approaches...

- Feature-based → *Syntax, Sentiment, Grammatical, and Readability* [3]
- Knowledge Bases
  - in Theory → ideal system: comparing article KB to reference KB [10]
  - in Practice → KG classification [6], DEAP-FAKED [9], TransE [2]
- Neural Networks → Text-transformers + ensemble with pseudo-labeling [8]

# Related Work

## A variety of approaches...

- Feature-based → *Syntax, Sentiment, Grammatical, and Readability* [3]
- Knowledge Bases
  - in Theory → ideal system: comparing article KB to reference KB [10]
  - in Practice → KG classification [6], DEAP-FAKED [9], TransE [2]
- Neural Networks → Text-transformers + ensemble with pseudo-labeling [8] and more...

# Table of Contents

1 Introduction

2 Related Work

3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

4 The App

5 Results

6 Conclusion

## LIAR-PLUS dataset

- **Training set** 10,269
- **Validation set** 1,284
- **Testing set** 1,283
- **Main topic** politics (economy, healthcare, taxes, federal budget, education, jobs, state budget, candidates biography, elections, and immigration)
- **Average sentence length** 20
- Similar vocabulary for all classes
- **Classes**
  - half-true: 2114
  - **false**: 1995
  - mostly-true: 1962
  - **true**: 1676
  - barely-true: 1654
  - pants-fire: 839

## Implemented fake news features

- Spelling mistakes
- Lexical repetitions
- Plural forms
- Loaded language
- Excessive usage of exclamation marks and imperative mood
- Punctuation errors
- References to past events

# Feature-based

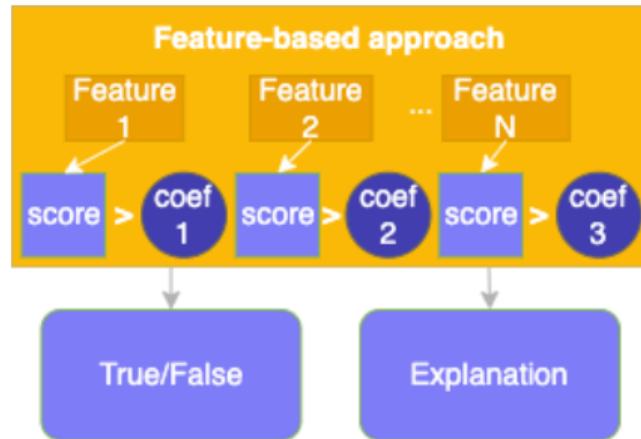


Figure: Feature-based module

# Cross-Checking

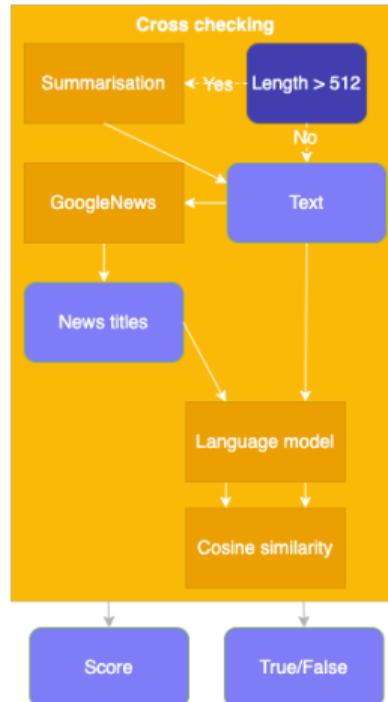


Figure: Cross-Checking module

# Neural Networks

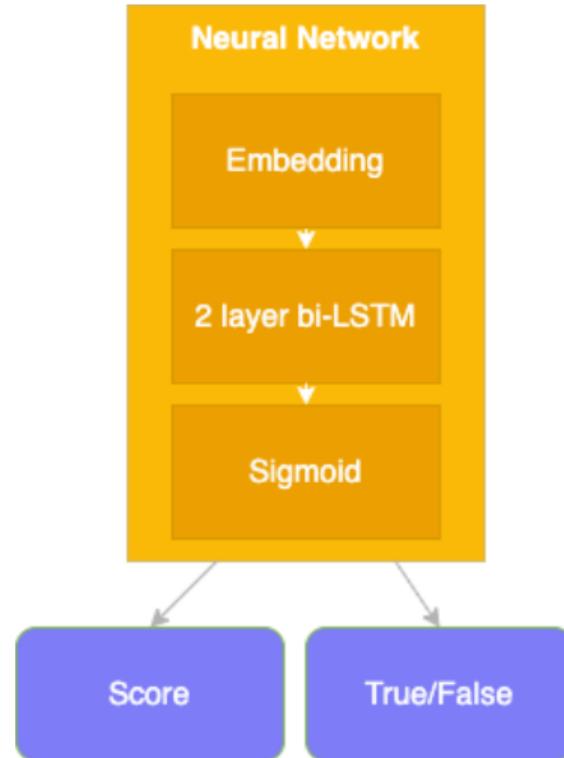


Figure: Neural Network module

## Options

- ✗ Ideal system → not enough data in reference KBs [10]
- ✓ Class frequency comparison<sup>1</sup>

---

<sup>1</sup>Solution proposed by *Mathieu D'Aquin, LORIA*

# Knowledge Base

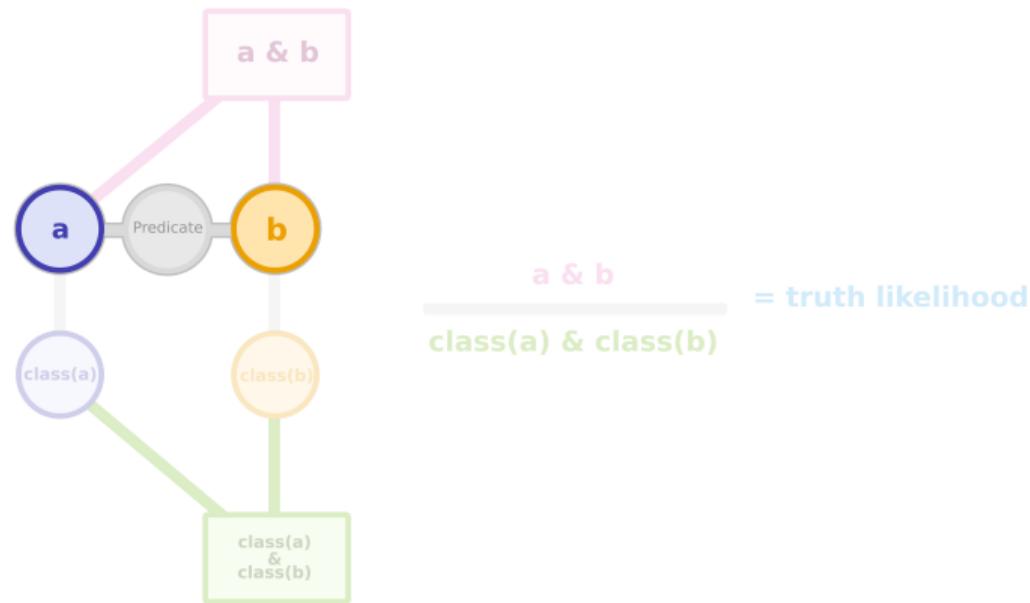


Figure: KB Pipeline: (1) extract a triple [...]

# Knowledge Base

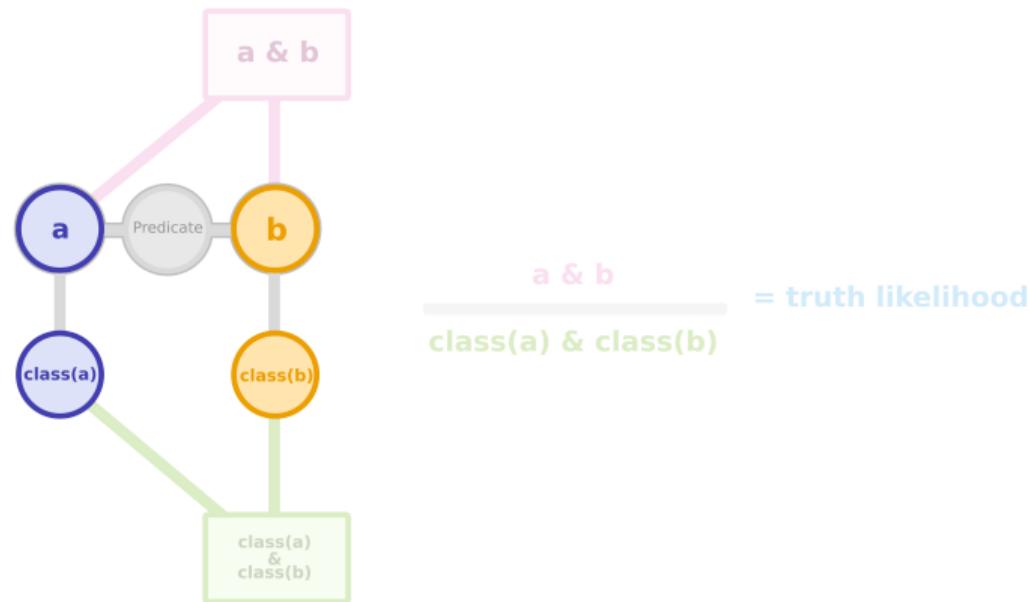


Figure: KB Pipeline: (1) extract a triple; (2) find the classes of subject  $a$  and object  $b$  [...]

# Knowledge Base

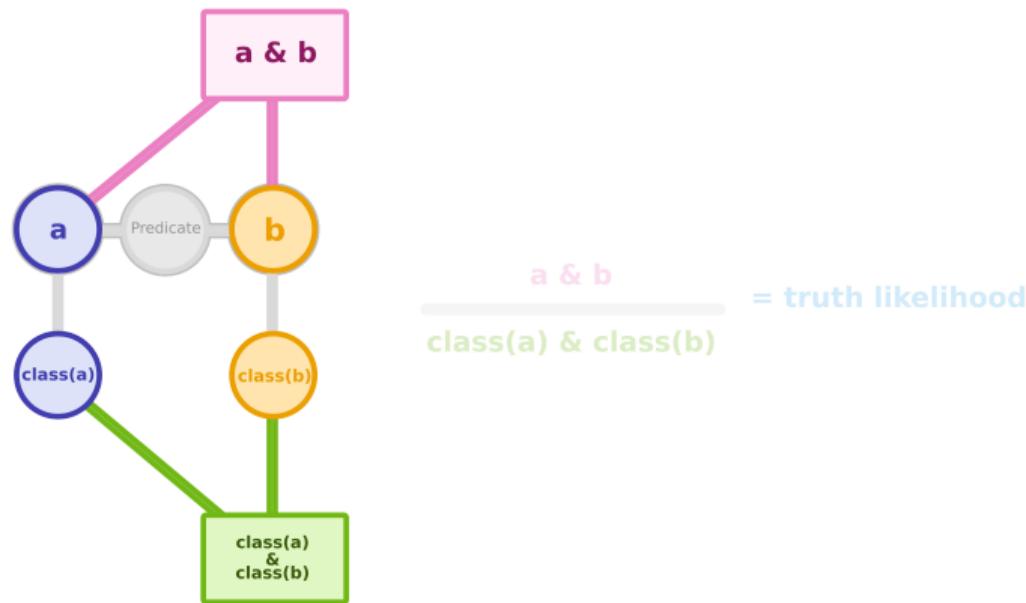


Figure: KB Pipeline: (1) extract a triple; (2) find the classes of subject *a* and object *b*; (3) count instances of triples containing *a* and *b* vs. those containing any member of such *class(a)* with any member of *class(b)* in a corpus [...]

# Knowledge Base

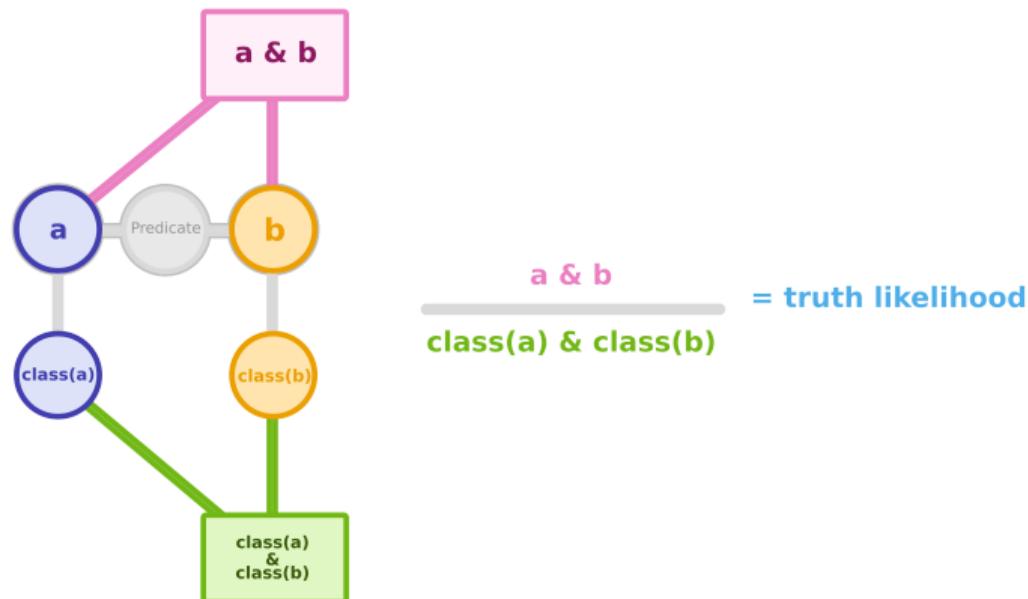


Figure: KB Pipeline: (1) extract a triple; (2) find the classes of subject  $a$  and object  $b$ ; (3) count instances of triples containing  $a$  and  $b$  vs. those containing any member of such  $class(a)$  with any member of  $class(b)$  in a corpus; (4) calculate the ratio.

# Knowledge base

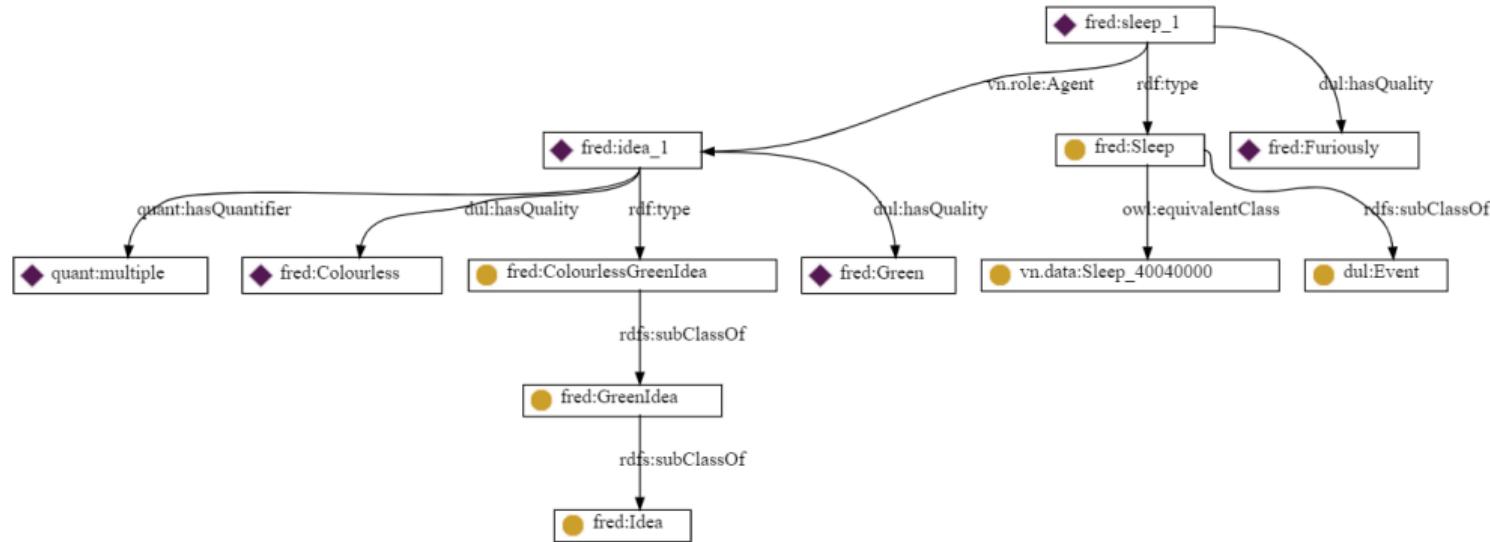


Figure: Example output of text to RDF conversion using FRED. [5]

# Ensemble

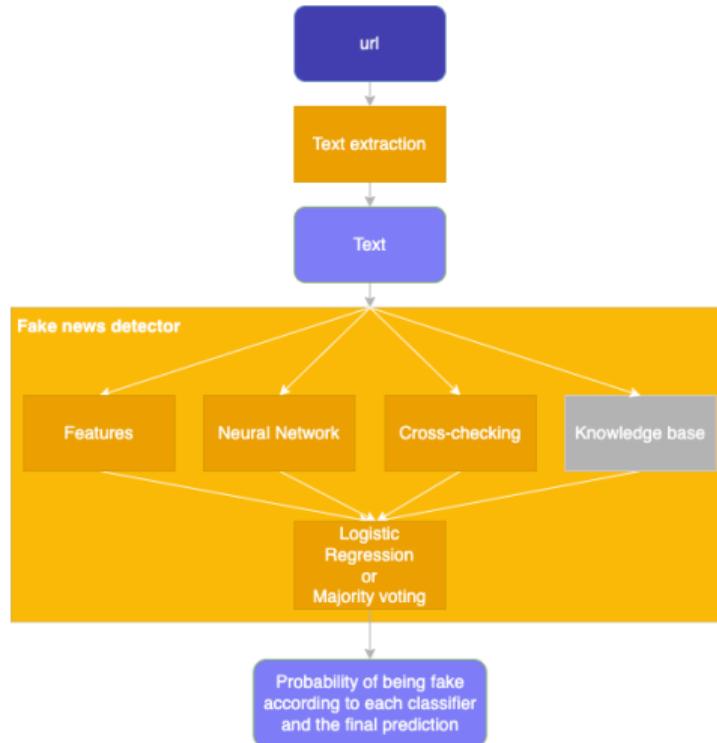


Figure: Modules ensemble

# Table of Contents

1 Introduction

2 Related Work

3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

4 The App

5 Results

6 Conclusion

## GUI options

- ✗ Extension → browser-dependent restrictions
- ✓ Web-app → ideal multi-platform compatibility

# GUI Structure

## Website - Homepage

The screenshot shows a dark-themed website homepage. At the top left is a logo with the letters 'JAKS'. To its right are 'Features' and 'About Us' links. A search bar with a magnifying glass icon and a 'Search' button are positioned at the top right. The main content area features a large, stylized graphic of a tree or plant on the left. Overlaid on the tree are several text snippets and a yellow call-to-action box. The text snippets include: 'Get your facts straight!', 'Try our Online App for Knowledge Substantiation', 'Check an article', and various event details like 'Eugene Net to Business Net working group, 11:00am' and 'Springfield 2nd Friday Art Walk 6pm-9pm, downtown library'. The overall aesthetic is modern and informative.

Are you getting real information?

Check how likely you are to be reading fake news. Are the sources credible? Is it a piece of propaganda? Are the claims founded? Discover in one click!



Have a look at some advanced insights

Take a peek at how our algorithm works and why it thinks your news sources are lying! Feel free to check all the stats to better understand our results.



Our Team



Elisa Lubrini

Research Intern @ Plant Health Institute of Montpellier



Anna Mosolova

Computational linguist, Master's student at Université de Lorraine



# GUI Structure

The screenshot shows the homepage of a website called FACTS. The header features a logo with three stylized letters 'F', 'A', and 'K' in white, followed by the word 'FACTS' in a bold, sans-serif font. Below the header, there are navigation links for 'Features' and 'About Us'. The main content area has a dark background with a faint, repeating pattern of small white shapes. A large yellow button at the top left contains the text 'Get your facts straight!'. Below it, a sub-headline reads 'Try our Online App for Knowledge Substantiation'. A yellow button labeled 'Check an article' is positioned below the headline. To the right, there is a section with several small, overlapping text boxes containing event details like 'Management Commemoration, 7-10am', 'Springfield Library, 225 Fifth St., Springfield, FREE', and 'Green Drinks, gathering of progressives, 5-7pm, New Day Cafe, 4409 Blair Blvd., FREE'. At the bottom, there is a form with a placeholder 'Enter URL.' and a blue 'Check' button.

This screenshot shows a detailed article page from the FACTS website. The header includes the 'FACTS' logo and a 'Go Back' link. The main content area displays a single article with a dark background and white text. The article's title is 'The likelihood of this article to be fake is of 3.14%'. Below the title, there is a short paragraph: 'More features are coming, such as statistics and algorithm insights... Stay tuned!' To the right of the text, there is a decorative graphic featuring two stylized figures in purple and blue, surrounded by various icons related to data analysis and technology.

The likelihood of this article  
to be fake is of 3.14%

More features are coming, such as statistics and algorithm insights...  
Stay tuned!

# Table of Contents

1 Introduction

2 Related Work

3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

4 The App

5 Results

6 Conclusion

# Results

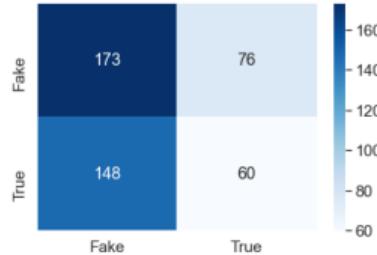


Figure: Feature-based module results

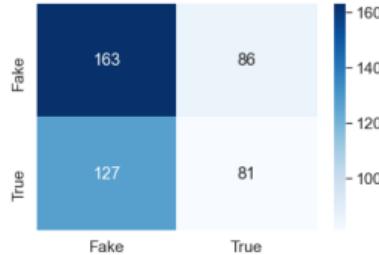


Figure: Cross-checking module results

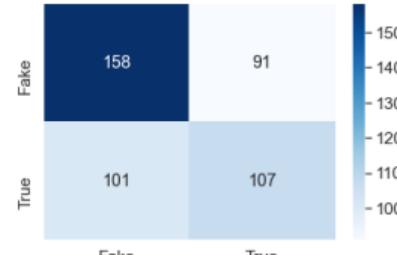


Figure: Neural network module results

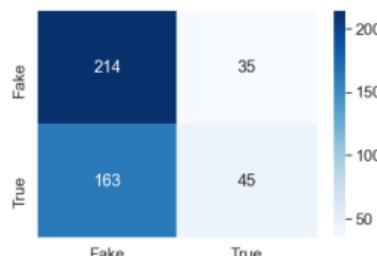


Figure: Logistic Regression ensemble results

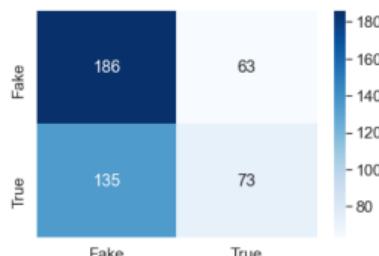


Figure: Voting ensemble results

# Table of Contents

1 Introduction

2 Related Work

3 Methodology

Dataset

Feature-based Module

Cross-Checking Module

Neural Networks Module

Knowledge Base Module

Ensemble

4 The App

5 Results

6 Conclusion

# Future Work

## Improving the app...

- Dataset
  - Fake News dataset with full articles<sup>2</sup>
- Results Displayed
  - Sentence-specific insights
- KB Module
  - Integrate the module in the pipeline
  - Let the user choose the modules
- Interactive Insights
  - Make the user experience smoother
- Loading Page
  - Discourse Analysis [1], Propagation [4], User based [7],...
- New modules

---

<sup>2</sup><https://www.kaggle.com/c/fake-news/overview>

thank  
you

