

Fake News Detection

Tuesday, April 8th 2025

CHIKHI Wassim

HOUADRIA Dounia

KHOUALDIA Besma

LIN Laurent

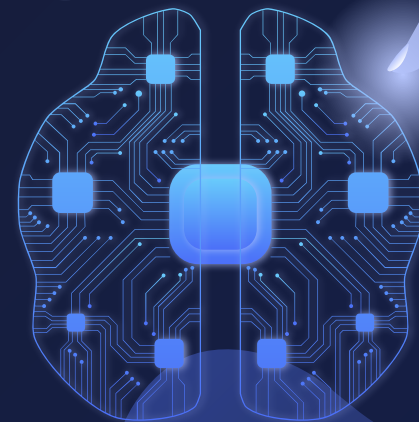




Table of contents

01

Introduction

04

Solution

02

Problems

05

Experimental
Results

03

Related work

06

Conclusion



Introduction



Fake News

A growing problem on social media.



Impact

Influences public opinion, elections,
economic crises.



Goals

Automatically detect fake news
using AI.



Problem Definition

What is fake news?

A fake news is an article containing misleading or false information.

How to detect them ?

- Which algorithm to use ?
- How to store data ?





Related Work



Traditional
methods

Manual fact-checking



Classical
approach

Keyword-based rules



Recent
advances

Deep Learning models (BERT, transformers)



Social
Network

Integration of social network analysis to improve detection

Proposed Solution (1)

Supervised algorithms

- Decision Tree (DT)
- Gradient Boosting (GB)
- Logistic Regression (LR)
- Support Vector Machines (SVM)
- Naive Bayes Algorithm (NB)
- Random Forest (RF)

Proposed Solution (2)

Pre-processing with Apache Spark

Cleaning

- Convert to lower case
- Remove punctuations, stop words and links

Bag of Words

- Represent text by word frequency
- Use 1500 most frequent words for features

Results (1)

Logistic Regression	SVM	Gradient Boost	Decision Tree	Random Forest	Naive Bayes
95 %	96 %	95 %	95 %	85 %	90 %

Accuracy performance of the algorithm for Fake News.

Results (2)

Logistic Regression	SVM	Gradient Boost	Decision Tree
99 %	99 %	96 %	96 %

Accuracy performance of the algorithm for True.csv+Fake.csv news.

Results (3)

Logistic Regression	SVM	Naive Bayes
95 %	92 %	87 %

Accuracy performance of the algorithm for Merged News dataset.

Our experiment (1)

FEVER dataset (22584 features post processing), 263 822 rows, 70066 Refutes, 193756 Supports

	Logistic Regression	SVM	Naive Bayes
Bag of Words (same as paper)	83.02%	82.92%	81.75%
TF-IDF (5000 features)	85.76%	85.68%	82.31%
Bag of Words (5000 features)	85.63%	85.44%	83.02%

Our experiment (2)

FEVER dataset but balanced dataset : 70066 Refutes, 70066 Supports

	Logistic Regression	SVM	Naive Bayes
Bag of Words (same as paper)	76.08%	75.92%	73.40%
TF-IDF (5000 features)	81.52%	81.71%	76.36%
Bag of Words (5000 features)	81.37%	81.58%	76.47%

Limitations and areas for improvement

Reproducibility

- No code provided
- Hyperparameter not given
- Limited details on dataset

Experimentation concerns

- 99% accuracy is unrealistically high for fake news

Improvements

- Test on real-world with evolving fake news
- Toward a real-time big data integration



Conclusion

- Fake news is a major challenge on social media.
- Classical machine learning methods perform well
- Spark makes big data handling smooth and efficient.
- Our experiments show good realistic results
- Despite strong results, real-world testing is needed



Thanks For Listening!

Research paper : B. Subaşı, H. Beral, N. Güleç, and T. Dökeroğlu, "Detecting Fake News on Big Data", Researcher, vol. 01, no. 02, pp. 1–5, 2021.

FEVER dataset : <https://huggingface.co/datasets/fever/fever>