

# Fake news detection

Simbiat Musa    Georg F.K. Höhn

Ironhack

19 June 2025

# Overview

This project tackles the problem of identifying fake news headlines using Natural Language Processing(NLP). A labeled dataset was used to train multiple models including Logistic Regression, Random Forest, KNN and XGB with both TF-IDF and GloVe text representation.

Introduction

Data overview

Methodology

Brief word about transformer models

Training results

Conclusion

# Introduction

# Introduction

This project focuses on detecting fake news headlines by applying a full end-end machine learning pipeline. Using a labeled dataset of news headlines, we trained several classical machine learning models and also leveraged state of the art transformer models from hugging face. We evaluated and compared model performance and ultimately produced prediction files to classify unseen news headlines as real or fake. - 34152 rows of headlines annotated as fake news (0) or real news (1)

## Aims

Classify texts in a test set as fake or real

## Data overview

- ▶ fake news: 17572
- ▶ real news: 16580
- ▶ relatively balanced

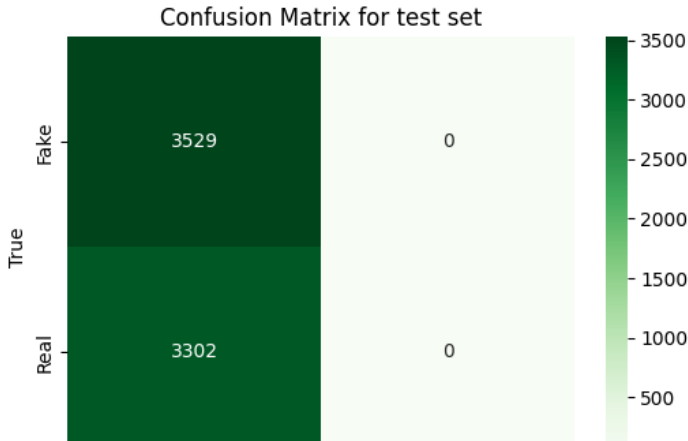
## Methodology





Brief word about transformer models

- ▶ tried pipelines with
  - 1) jy46604790/Fake-News-Bert-Detect
  - 2) omykhailiv/bert-fake-news-recognition
- ▶ both performed abysmally: everything is fake, see result for 1) below
- ▶ not sure why, over-sensitive? something wrong with our data pre-processing?



## Training results



## Conclusion



## Collaboration

- ▶ used py files to avoid notebook consistency issues with git
- ▶ VS Code offers the possibility of generating jupyter-like cells with # %%

Thanks for your attention!