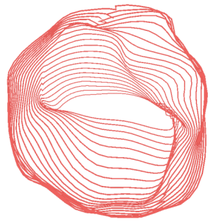
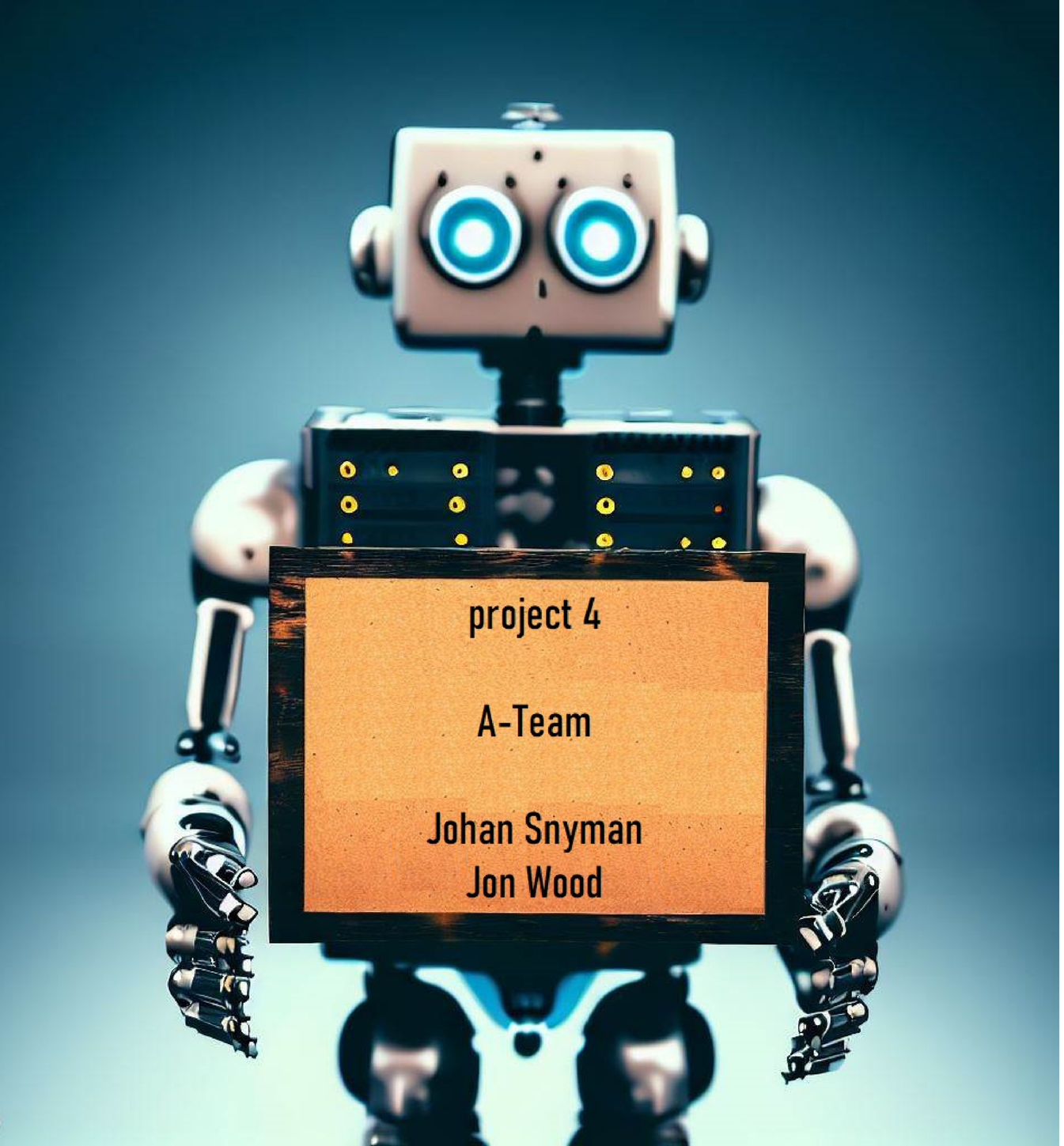
Detecting Fake News Using Machine Learning

PROJECT 4 PRESENTATION





Agenda

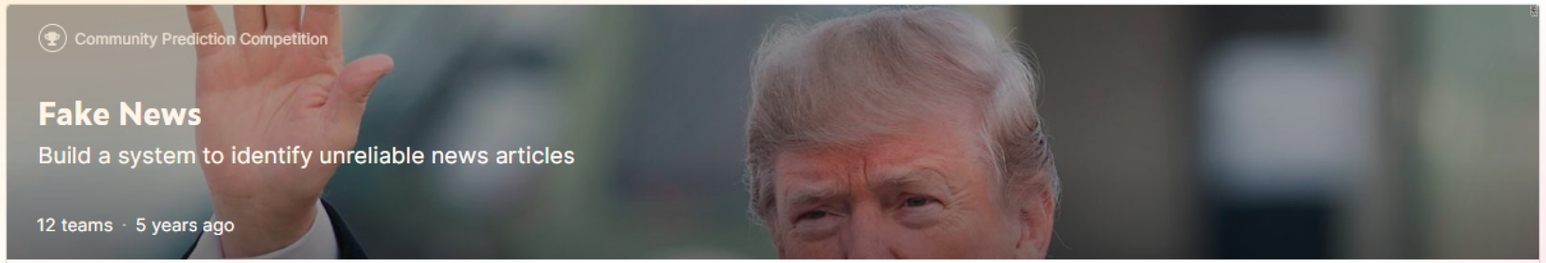
- Introduction
- Data Model Implementation
- ⚙ Data Model Optimisation
- GitHub Documentation
- Conclusion

Introduction

The pressing problem of fake news
Why detecting fake news is crucial
Machine Learning to the rescue
Technologies used in our project

DATASET

https://www.kaggle.com/competitions/fake-news/data



Overview | Data | Code | Discussion | Leaderboard | Rules | Team | Submissions | Late Submission | ...

Dataset Description

train.csv: A full training dataset with the following attributes:

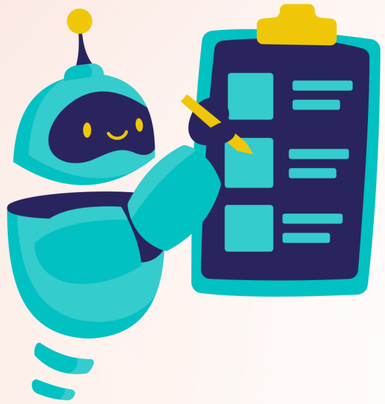
- **id**: unique id for a news article
- **title**: the title of a news article
- **author**: author of the news article
- **text**: the text of the article; could be incomplete
- **label**: a label that marks the article as potentially unreliable
  - 1: unreliable
  - 0: reliable

test.csv: A testing training dataset with all the same attributes at train.csv without the label.

Files
3 files

Size
123.81 MB

Type
csv



Data Model Implementation

- Python magic: Model initialization, training, and evaluation
- Data transformation: Cleaning, normalization, and standardization
- Power of SQL: Model utilizing data retrieved
- Impressive results: Model achieving 93% classification accuracy

Data Cleaning and Preprocessing

Steps taken for data cleaning

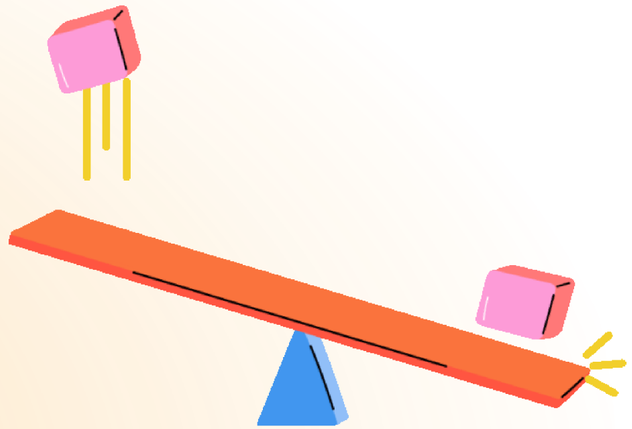
Normalization and standardization techniques used

Dataset: 26000+ Entries

Data loaded to PostgreSQL for analysis

Model Selection and Performance

Model choice and reasons
Scikit-learn for ML model implementation
Text preprocessing, feature extraction, and model selection
Model performance metrics: 95% accuracy on the test set



main · 5 branches · 0 tags · Go to file · Add file · Code

jonowood Update app.py — 8e253e6 1 hour ago — 62 commits

Dataset	Add files via upload	last week
ETL	Showing optimization	yesterday
ML	Update JONO_PREDICT_MASTER.ipynb	1 hour ago
Pickles	heroku requirements and code changes	19 hours ago
Presentation	start presentation	yesterday
Proposal	Create Project 4 - A-TEAM - Proposal.pdf	2 weeks ago
Static/Images	Branched	last week
nltk_data/corpora/stopwords	ffs	1 hour ago
static	ffs	1 hour ago
templates	Update index.html	17 hours ago
.gitignore	heroku	18 hours ago
Procfile	added stopwords locally	1 hour ago
README.md	Update README.md	3 weeks ago
app.py	Update app.py	1 hour ago
nltk.txt	Create nltk.txt	1 hour ago
requirements.txt	gfh	17 hours ago

README.md

About

UWA D
Submis

Rea
3 st
1 wa
1 fo

Releas

No releas
Create a

Packag

No packa
Publish y

Contrib

jo
Jo

Enviro

nlp

Langua

Jupy
Othe

GitHub Documentation

- Repository organization and .gitignore usage
- Customized README for polished presentation
- Project structure
- Overview of the repository structure
- Key contents and files
- Web application setup and usage

Conclusion

Recap of project accomplishments

Future improvements and extensions:
Advanced techniques, such as deep learning
Increasing the dataset size
Deploying the web application on a cloud platform

Acknowledgments

Thanks to our bootcamp instructors for their guidance and support

Questions & Discussions ?

Open for audience questions and lively discussion

Johan Snyman
JohanfromEsperance

THE END

Jono Wood
jonowood