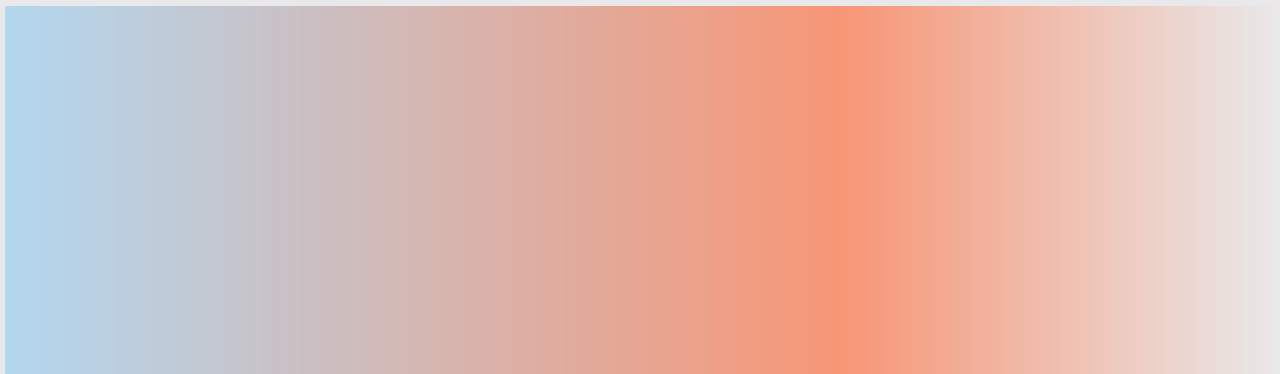




PROJECT 4 PROPOSAL



PREDICTION

ADAPTION

AUTOMATION

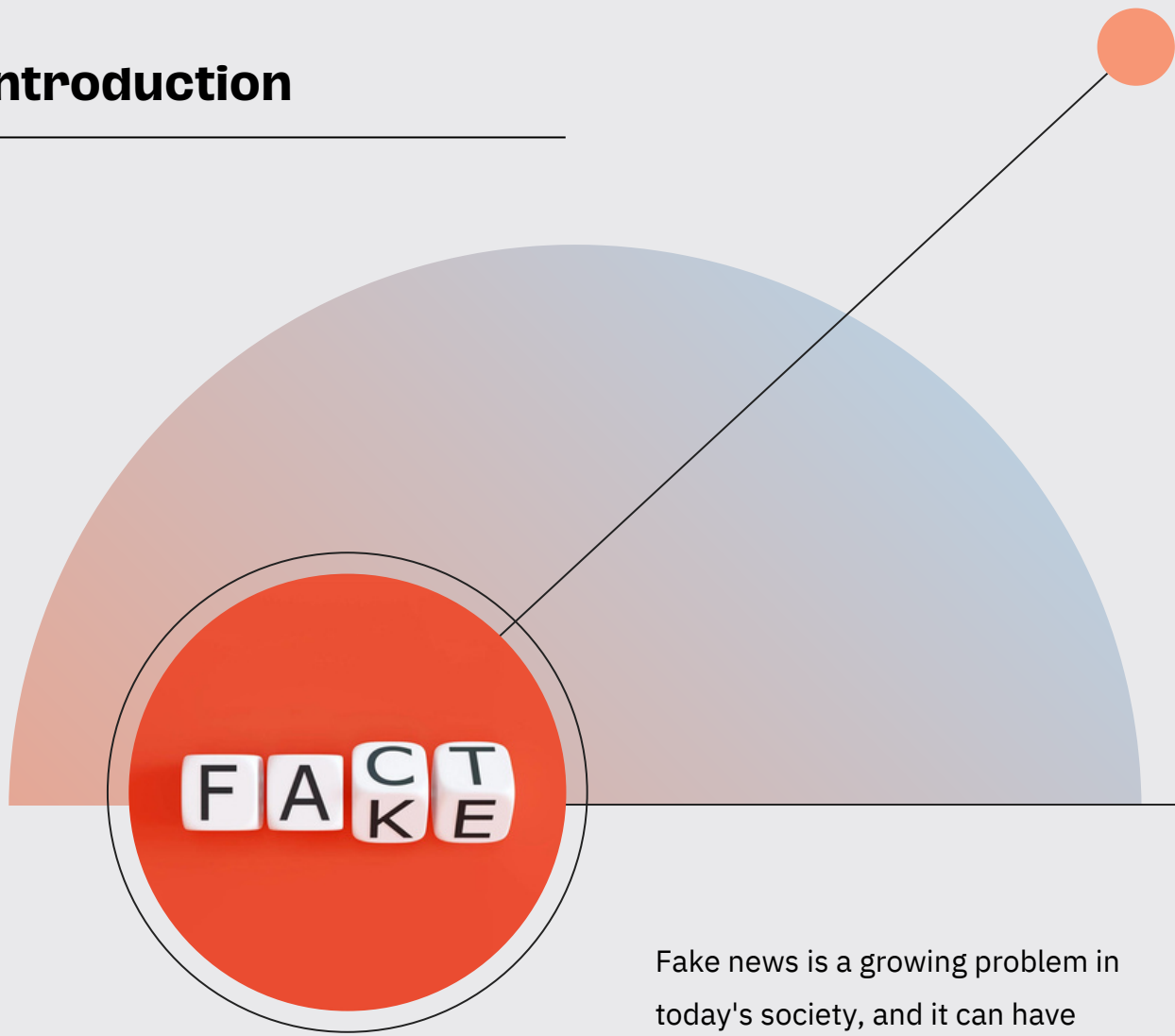
THE A TEAM



2023

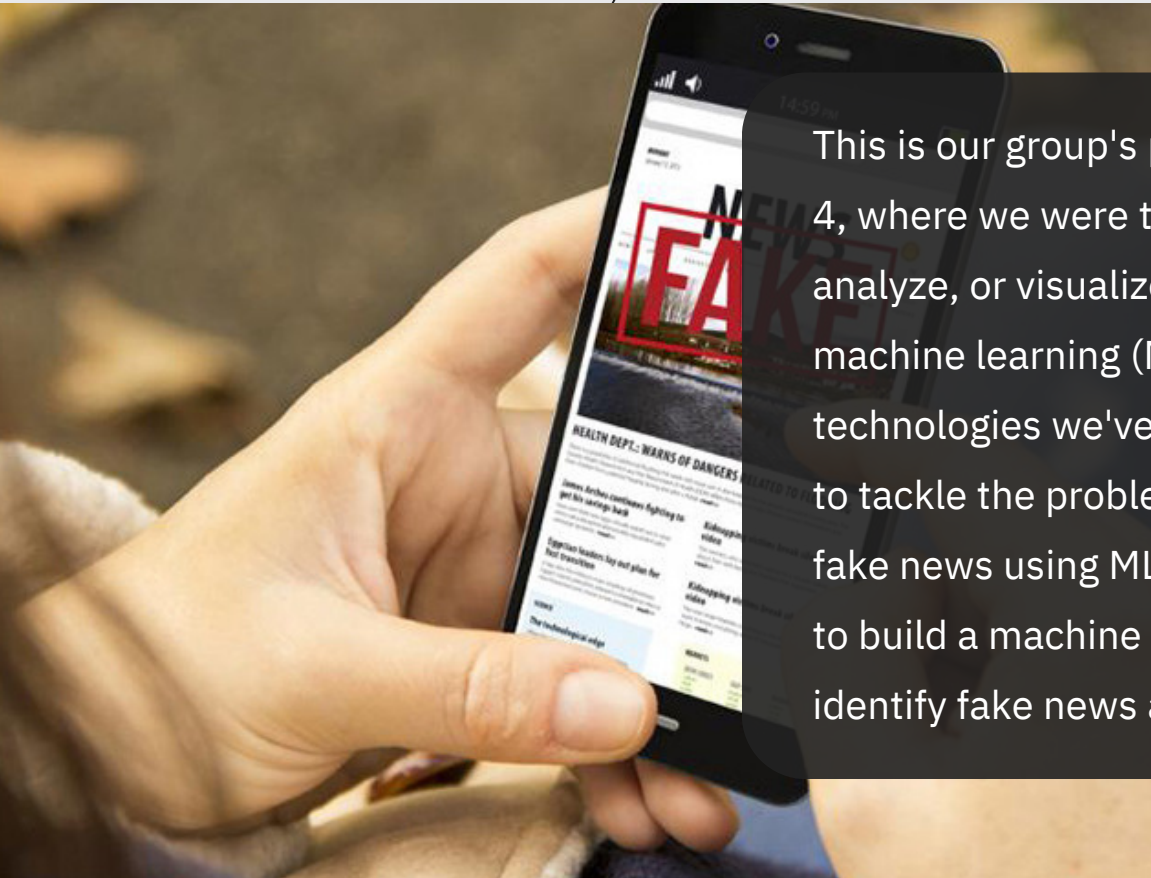
<https://github.com/jonowood/Project-4-A-Team>

Introduction



Fake news is a growing problem in today's society, and it can have serious consequences, including influencing public opinion and even shaping policy decisions. To address this problem, we propose to build a machine learning model that can accurately identify whether a given news article is fake or not. We aim to use various machine learning techniques such as text preprocessing, feature extraction, and model selection to achieve high accuracy.

Project Overview

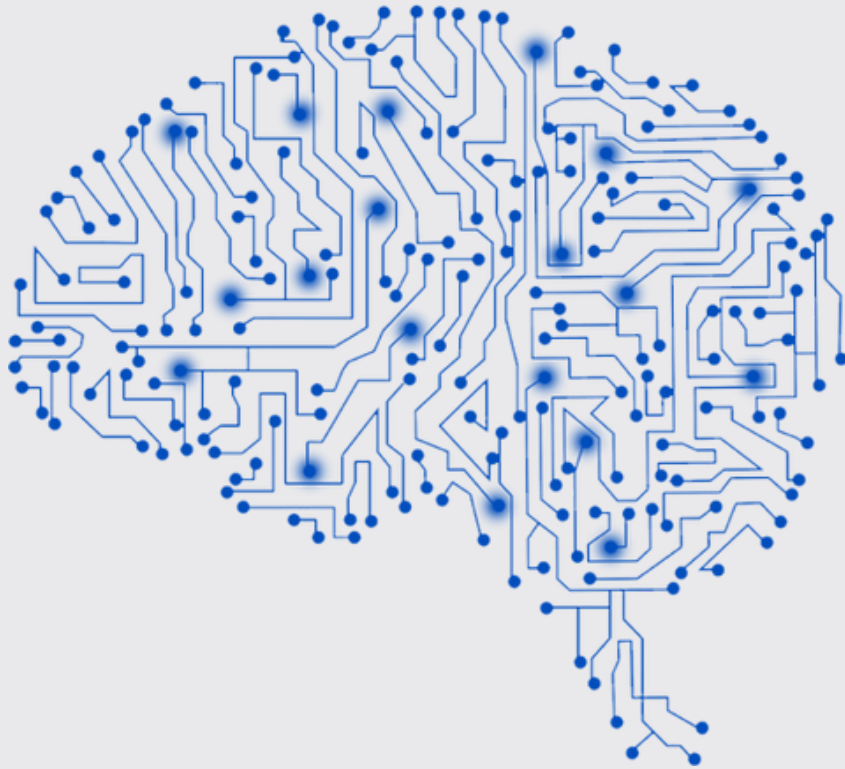


This is our group's project for Project 4, where we were tasked to solve, analyze, or visualize a problem using machine learning (ML) with the other technologies we've learned. We chose to tackle the problem of identifying fake news using ML. Our project aims to build a machine learning model to identify fake news articles accurately.

The project will involve the following steps:

1. **Data Collection:** Collect a dataset of news articles, consisting of both real and fake news articles.
2. **Data Preprocessing:** Preprocess the data to remove any irrelevant information, normalize and standardize the data, and split the dataset into training and testing sets.
3. **Feature Extraction:** Extract features from the news articles using various techniques, such as word embeddings, bag-of-words, and TF-IDF.
4. **Model Selection:** Experiment with different machine learning algorithms, such as logistic regression, decision trees, and neural networks, to select the best-performing model.
5. **Model Evaluation:** We will evaluate the performance of the selected model on the testing set using various metrics, such as accuracy, precision, recall, and F1-score.
6. **Web Application:** We will develop a web application that will allow users to input news articles and get the prediction of whether the article is fake or real.

Technologies



We will use the following technologies for this project:

1. Python: Used for data transformation.
2. Scikit-learn: Used for data preprocessing, feature extraction, and model selection.
3. Pandas: Used for data manipulation and analysis.
4. Matplotlib: Used for data visualization.
5. HTML/CSS/Bootstrap: We will use HTML, CSS, and Bootstrap for frontend web development.
6. Flask: Used for building web applications.
7. Postgres: We will use PostgreSQL for data storage and retrieval.



Data Set



We will use a publicly available dataset consisting of news articles, which have been labeled as real or fake. The dataset will be preprocessed to remove any irrelevant information.

This dataset will contain a minimum of 100 records, 500 fake news and 500 real news articles.

Model Building

We will build a machine learning model using Scikit-learn to classify news articles as real or fake. We will use various techniques such as text preprocessing, feature extraction, and model selection to achieve high accuracy. The model will be trained on the training set and tested on the testing set. We will evaluate the model's performance using various metrics such as accuracy, precision, recall, and F1-score.

Conclusion

Our project aims to build a machine learning model to identify fake news articles accurately. The model will be built using various techniques such as text preprocessing, feature extraction, and model selection. We will evaluate the model's performance using various metrics such as accuracy, precision, recall, and F1-score. We will also develop a web application that will allow users to input news articles and get the prediction of whether the article is fake or real.

Our Team



Johan Snyman

<https://github.com/JohanfromEsperance>



Jon Wood

<https://github.com/jonowood/>

<https://github.com/jonowood/Project-4-A-Team>



THANK YOU

PREDICTION
ADAPTION
AUTOMATION

THE A TEAM

<https://github.com/jonowood/Project-4-A-Team>