Fake News Detector

Using ML to Help Tackle Misinformation

Jesús Badenes @ Ironhack

What will we cover today?

01

Why Solve for This?

02

The Approach

03

The Data

04

The Models

05

The Application

06

Final Reflections

01

Why Solve for This?

The State of Affairs

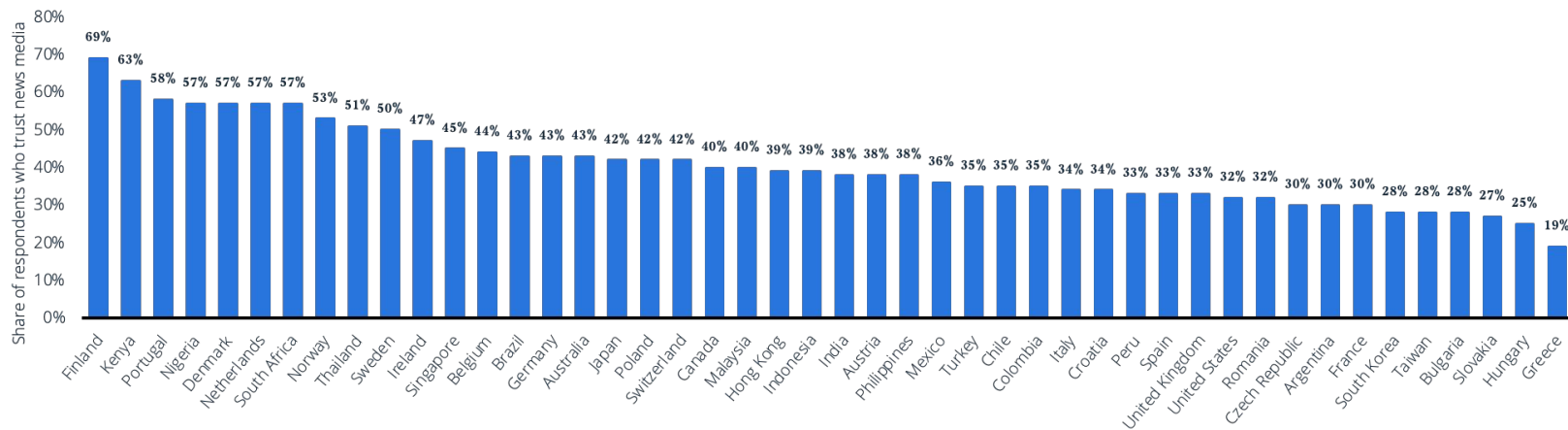### Creating and Distributing Fake News Has Never Been Easier

- **AI** tools produce text **massively**.
- **Social media** amplifies **misinformation** and is a growing source.

Trust in Media Has Never Been Lower

- Ipsos* reports a **5% global drop** in **media trust** over 5Y.
- Reasons: **fake news** spread and doubts about **media integrity**.

Good Journalism is at the Heart of Democracy

- Research links **healthy media** to **effective democracies**.
- **Media trust** can boost **voter turnout** by **13%**.

Share of respondents who trust news media

Country	%
Finland	69%
Kenya	63%
Portugal	58%
Nigeria	57%
Denmark	57%
Netherlands	57%
South Africa	57%
Norway	53%
Thailand	51%
Sweden	50%
Ireland	47%
Singapore	45%
Belgium	44%
Brazil	43%
Germany	43%
Australia	43%
Japan	42%
Poland	42%
Switzerland	42%
Canada	40%
Malaysia	40%
Hong Kong	39%
Indonesia	39%
India	38%
Austria	38%
Philippines	38%
Mexico	36%
Turkey	35%
Chile	35%
Colombia	35%
Italy	34%
Croatia	34%
Peru	33%
Spain	33%
United Kingdom	33%
United States	32%
Romania	32%
Czech Republic	30%
Argentina	30%
France	30%
South Korea	28%
Taiwan	28%
Bulgaria	28%
Slovakia	27%
Hungary	25%
Greece	19%

Why is Identifying Fake News so Difficult?

### Information x 10^n = No Information

- The **Internet** offers **vast information**, but humans struggle to process it all.
- Selecting reliable information is **effort-intensive** and **time-consuming**.

Algorithmic Echo Chambers
- We manage information through **quality news** (often costly) and **algorithmic curation**.
- Social network algorithms often prioritize **attention over truth**, showing us preferred (even if untrue) content.

Confirmation Bias

- People naturally favor information that **aligns with their beliefs**.
- This **"Confirmation Bias"** makes detecting fake news hard.

OBJECTIVE FACTS

WHAT CONFIRMS YOUR BELIEFS

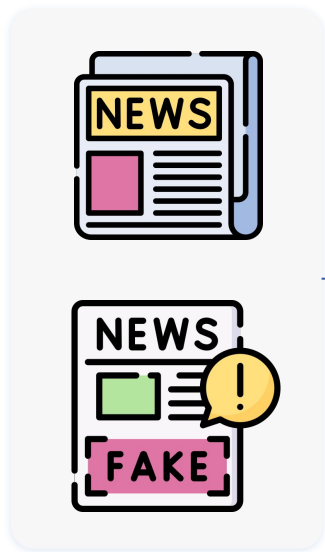WHAT YOU SEE
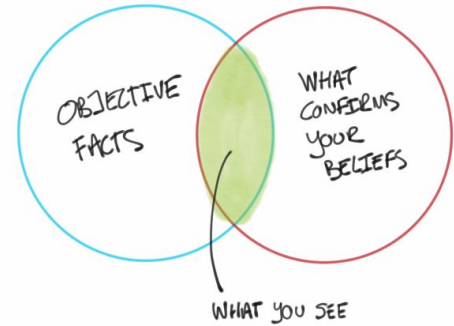
02

The Approach

Identifying Fake News at Scale Through ML

Using a trained model, we can quickly help identify potential fake news

Dataset of Fake and True News

Binary Classification Algorithms

User Application

Data Cleanup

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine
- Feedforward Neural Network
- Recurrent Neural Network

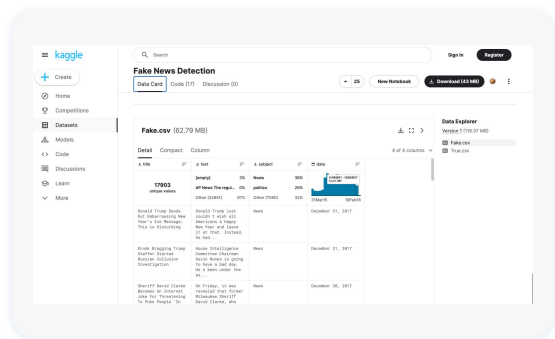Model Deployment

Deploy Model for User Application

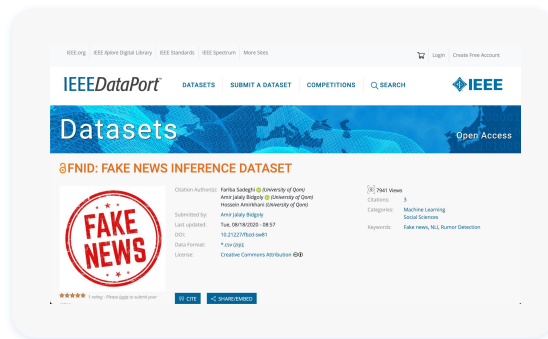Streamlit

03

The Data

Data Sources

To train a binary classification model, we need a labeled & diverse dataset of news articles

Kaggle Dataset (40k)

IEEE Dataset (20k)

Modeled Features

❖ News Title

❖ News Text

Challenges:

- It is essential to train our models on previously classed data from fact checking organizations.
- We expanded our dataset to two sources due to overfitting concerns.

Data Cleaning

Removing noise boosts model accuracy, ensures consistency, minimizes sparsity for better efficiency, and helps in reducing biases, ultimately promoting better adaptability to new data
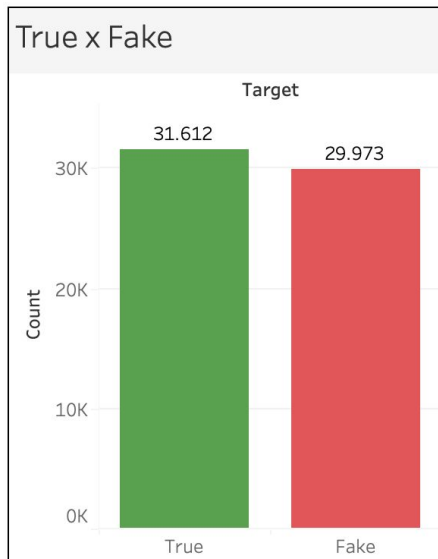
What Key Actions Did We Take?

- **Lowercasing**: Convert all text characters to **lowercase** for consistency.
- **Removing Brackets**: Eliminate **square brackets** and their enclosed contents.
- **Non-Word Characters**: Replace **non-word** characters with spaces.
- **URL Removal**: Strip out **URLs** and web addresses.
- **HTML Tag Removal**: Erase **HTML tags** from the text.
- **Punctuation Removal**: Discard **punctuation** marks for cleaner text.
- **Newline Character Removal**: Erase **newline characters** for continuous text.
- **Alphanumeric Word Removal**: Strip words containing **digits**.
- **Expanding Contractions**: Convert **contractions** (e.g., "it's" to "it is").
- **Stopword Removal & Lemmatization**: Remove **common words** (e.g., "is", "an") and reduce words to their root form using **lemmatization**.
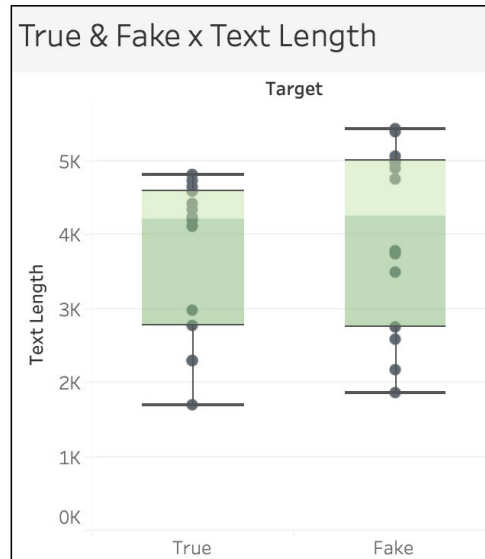
Exploratory Data Analysis

Understanding the distribution of our data and analyzing patterns

True x Fake

Target

31.612

29.973

30K

20K

Count

10K

0K

True   Fake

True & Fake x Text Length

Target

5K

4K

Text Length

3K
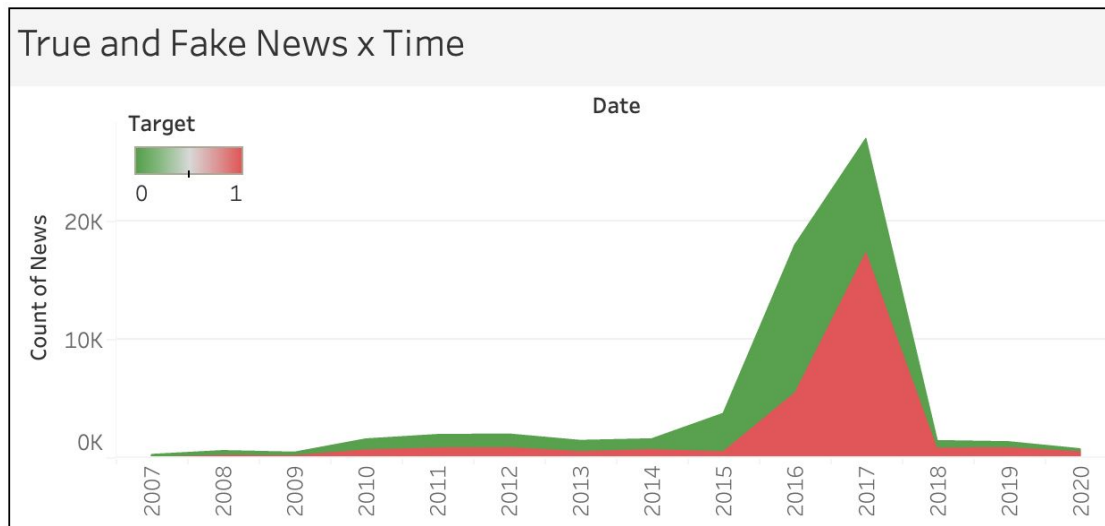
2K

1K

0K

True   Fake

Worked with balanced data

Observed that fake news were slightly longer

Exploratory Data Analysis

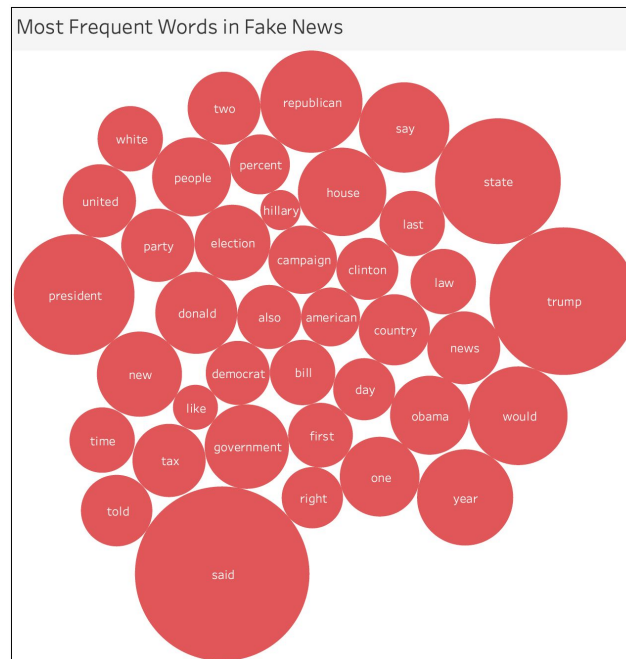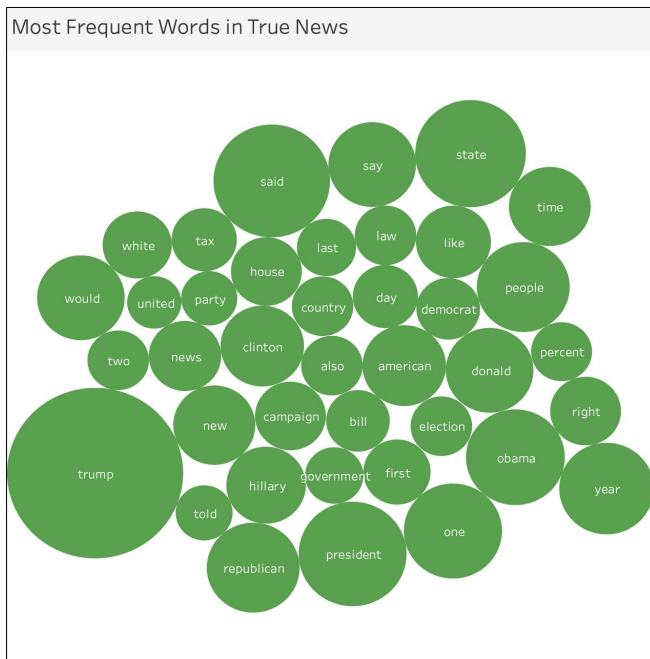Visualising a time series of news articles in the training set

A majority of the articles in the training set are from 2015 – 2018, and they are all in the range of 2007 – 2020

Exploratory Data Analysis (3 / 3)

Visualising most common words across true and fake news

We can observe a significant overlap in most common words, but there are key differences

04

The Models

The Models

After cleaning the data, we ran 6 models well fit for classification tasks

Model Results

Model	Accuracy
Logistic Regression	0.89
Decision Tree Classifier	0.85
Random Forest Classifier	0.88
Support Vector Machine	0.89
Feedforward Neural Net	0.84
Recurrent Neural Net	**0.90**

Key Observations

RNN Excellence: With a 0.90 accuracy, the Recurrent Neural Net outperforms all, showing its strength in capturing sequential patterns.

Classic Algorithms Shine: Logistic Regression and SVM both achieve an impressive 0.89 accuracy, highlighting their competitive performance.

Robust Data: Most models hover around upper 80% accuracy, indicating consistent and clean data.

Ensemble Benefit: Random Forest (0.88) outperforms its single counterpart, the Decision Tree (0.85), underscoring ensemble methods' advantage.
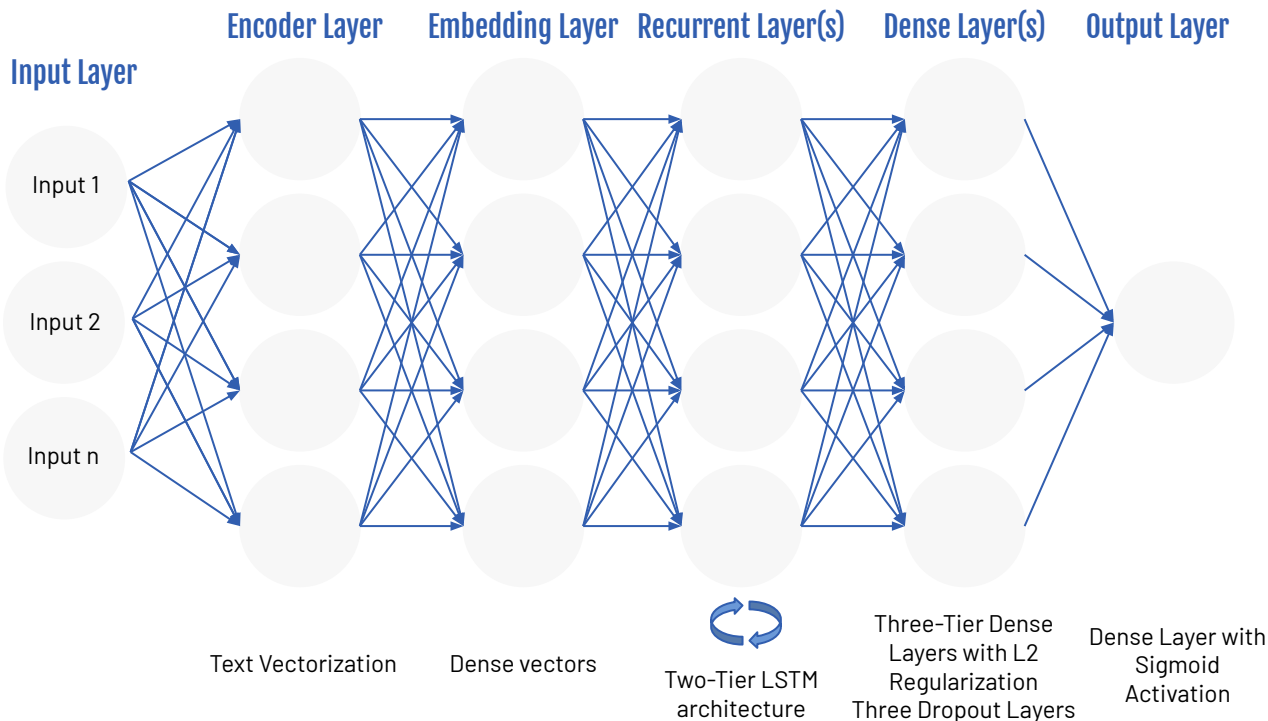
Simplicity vs. Structure: The RNN's superior performance over the Feedforward Neural Net (0.90 vs. 0.84) suggests the importance of data structure.

The Selected Model: Recurrent Neural Network

Our final model is a recurrent neural network with an 11-layer architecture

Encoder Layer **Embedding Layer** **Recurrent Layer(s)** **Dense Layer(s)** **Output Layer**

Input Layer

Input 1

Input 2

Input n

Text Vectorization

Dense vectors

Two-Tier LSTM architecture

Three-Tier Dense Layers with L2 Regularization Three Dropout Layers

Dense Layer with Sigmoid Activation

Challenges:
- Text length and vocabulary size
- Overfitting & lack of generalization

05

The Application

Empowering Users to Identify Fake News

Creating a Streamlit App for users to paste news articles and assess their credibility

Predicting True News...

Fake News Detector

News Title

President Donald Trump questioned the politics c

News Text

President Donald Trump questioned the politics of his decision to finally acknowledge former President Barack Obama was born in the US, which he did late during the campaign in 2016, according to a source close to the White House.

The source said that shortly after he made the statement, Trump told aides that he would have done better in the polls had he continued to stand his ground on the birth certificate issue.

Trump has continued to question the legitimacy of Obama's birth certificate during private conversations in recent months, The New York Times reported Tuesday, citing

Predict

This news is likely true.

The model predicts a probability of `47.87` % that this news is fake.

Disclaimer: This prediction is based on a Deep Learning model called LSTM (Long Short-Term Memory). It is based on a variety of recurrent neural networks (RNNs) that are capable of learning long-term dependencies, especially in sequence prediction problems. This is a research project and as such, it may be prone to errors. We recommend doing further research before determining whether a piece of news is true or fake.

... And Fake News

Fake News Detector

News Title

The Pope's Coat Is Here to Ruin Your Faith

News Text

Future generations will acknowledge the vibe shift. It happened last weekend, when all of a sudden social media feeds filled with images of Pope Francis, typically a pious and plain dude, looking like a boss in a sleek white puffer coat. It was instantly a meme, a LOL in a sea of bad news. It was also not real. Someone created the image using the artificial intelligence tool Midjourney. But it fooled a lot of people—so many that news outlets began calling it "one of the first instances of wide-scale misinformation stemming from artificial intelligence."

Just typing that sentence feels haunting. Like the first time you see someone in a red cloak in The Handmaid's Tale. Not that this

Predict

This news is likely fake.

The model predicts a probability of `87.34` % that this news is fake.

Disclaimer: This prediction is based on a Deep Learning model called LSTM (Long Short-Term Memory). It is based on a variety of recurrent neural networks (RNNs) that are capable of learning long-term dependencies, especially in sequence prediction problems. This is a research project and as such, it may be prone to errors. We recommend doing further research before determining whether a piece of news is true or fake.

06

Final Reflections

Challenges and Future Developments

Challenges

- **Data Diversity**: If the training data isn't diverse enough, the model might become too specialized, hindering its performance on unseen data.
- **Overfitting Risk**: While the model includes dropout and kernel regularization to combat overfitting, there's always a risk, especially with deeper networks.
- **Model Interpretability**: LSTMs, like other deep learning models, are often considered "black boxes", making it challenging to understand and explain their decision-making processes.
- **Scalability**: As the dataset grows, the computational cost of training such a complex model will increase.
- **Research & Confirmation Bias**: Relying exclusively on predictions to identify fake news can be risky and the need for the user to do further research is a limitation.

Future Developments

- **Expanding Data**: Adding more data sources will increase the model's ability to generalize.
- **BERT and Variants**: Leveraging models like BERT, RoBERTa, or DistilBERT which have been pre-trained on massive corpora can bring in the advantage of extensive prior knowledge.

Thank you