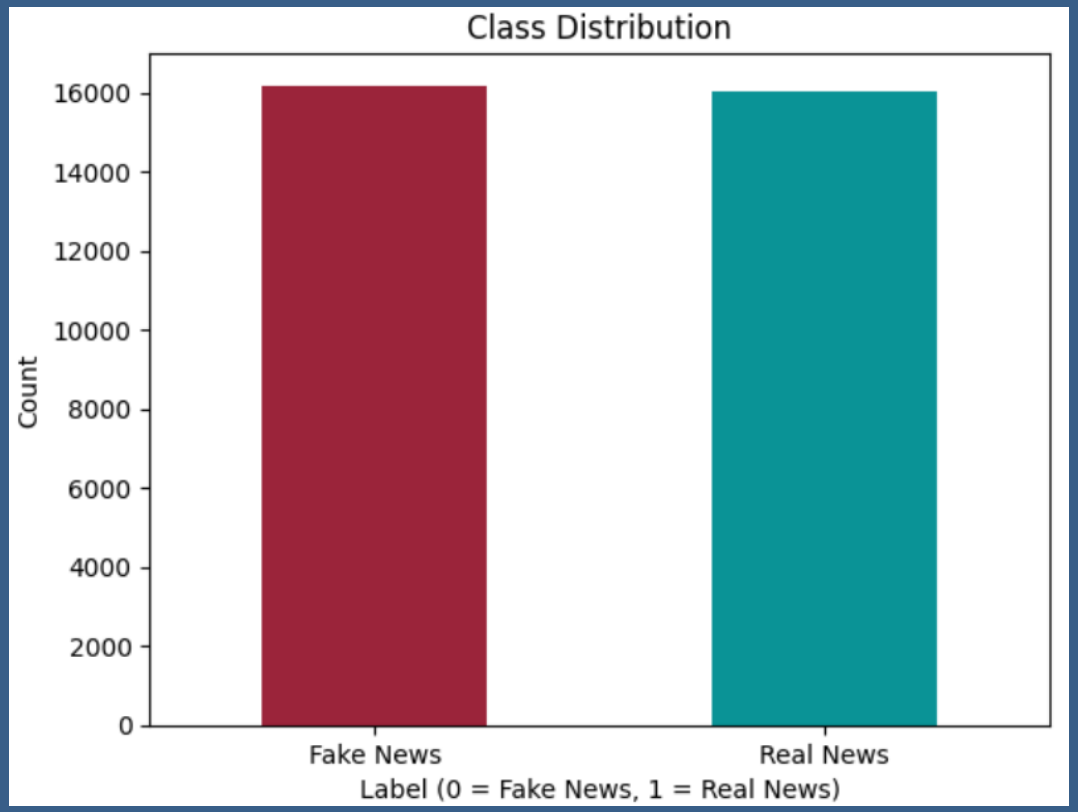
FAKE NEWS DETECTION

Ana
Morais

PROJECT OVERVIEW

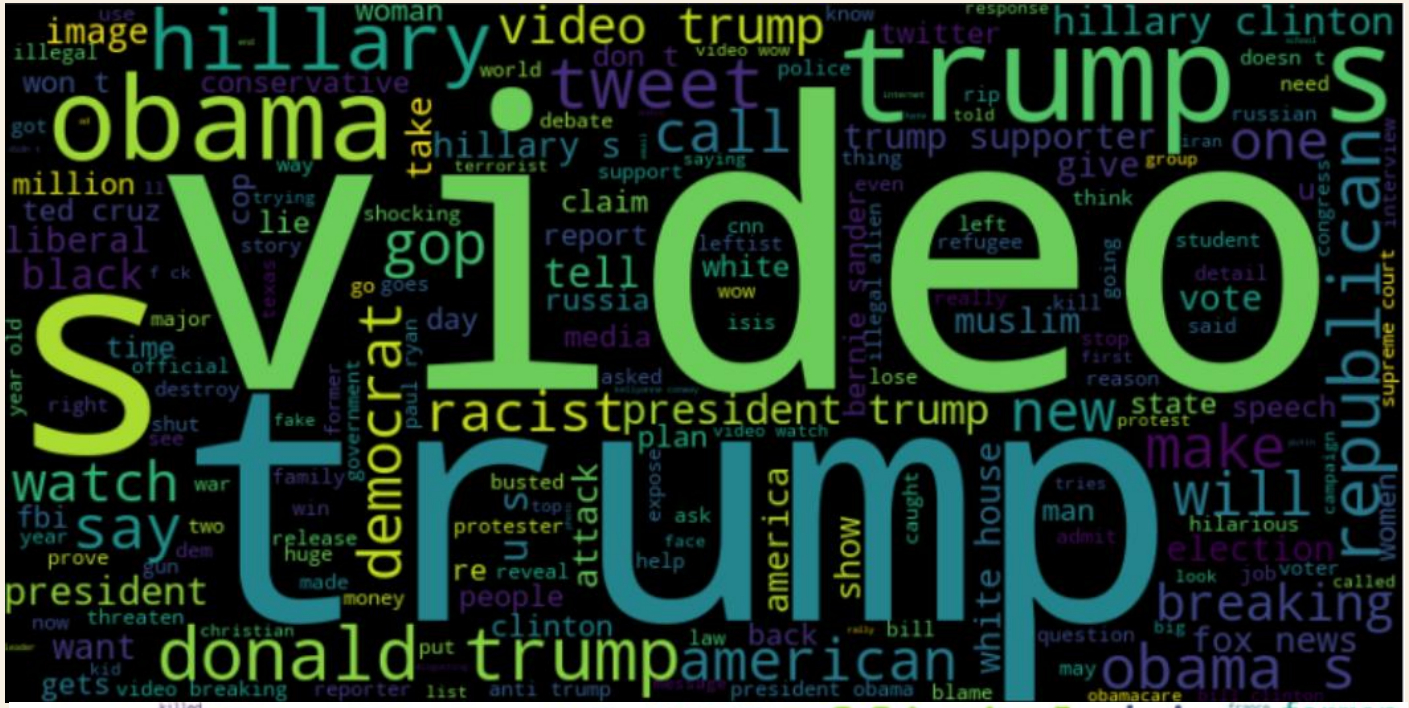
- **Objective**: Build a classifier to distinguish between real and fake news headlines.
- **Dataset**:
  - training_data.csv
  - contains news headlines and labels: 0 = Fake, 1 = Real)

BALANCED DATASET

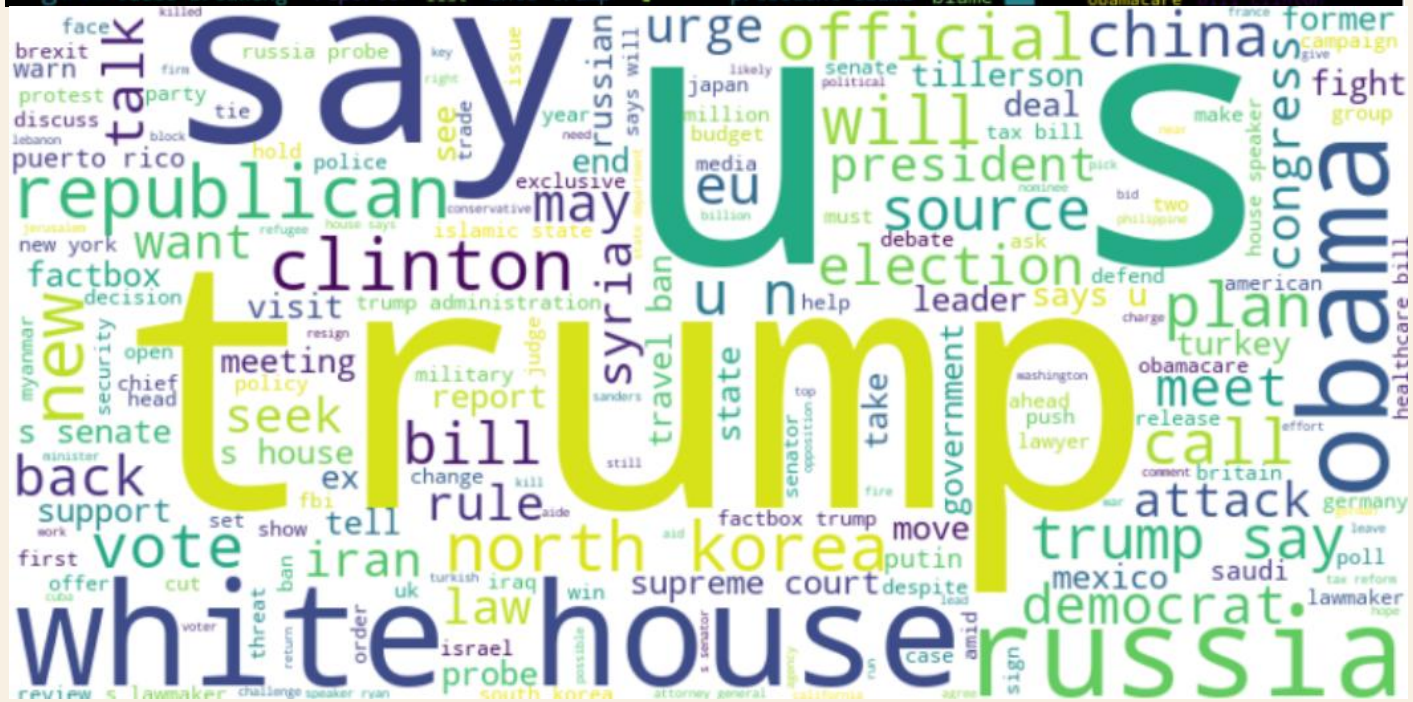


tag	counts
0	17572
1	16580
T	**34152**

Fake News



Real News



IMPLEMENTATION DETAILS (BASE MODEL)

Data preprocessing:

- Train-Test Split
- TF-IDF vectorizer

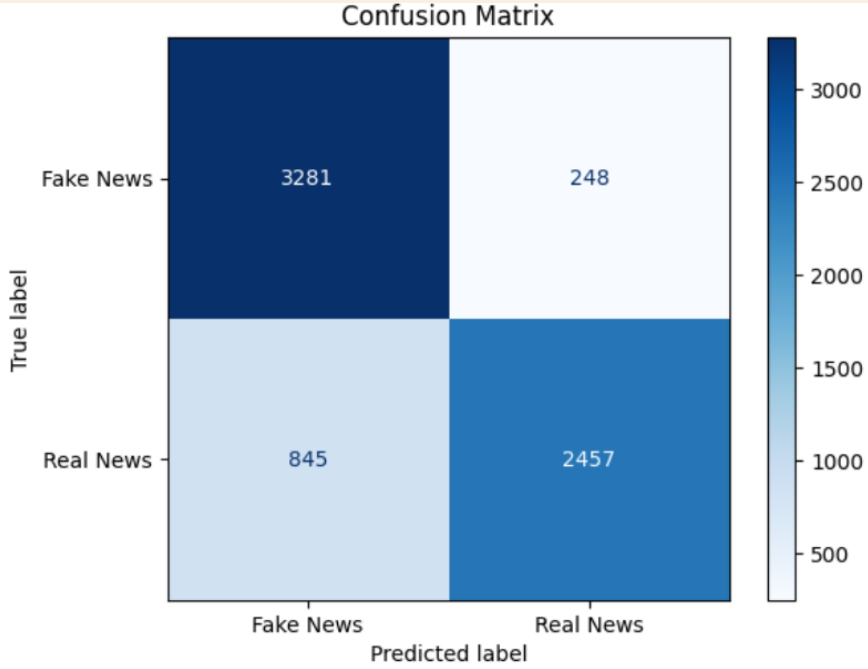
Base Model: Random Forest Classifier

- N_estimators=200
- Criterion='entropy'
- Max_depth=10

Evaluation Metrics:

- Accuracy, Precision, Recall, F1-Score

BASE MODEL EVALUATION



Classification Report:

	precision	recall	f1-score	support
0	0.80	0.93	0.86	3529
1	0.91	0.74	0.82	3302
accuracy			0.84	6831
macro avg	0.85	0.84	0.84	6831
weighted avg	0.85	0.84	0.84	6831

IMPLEMENTATION DETAILS (BASE MODEL + LEMMATIZATION)

Data preprocessing:

- Removed duplicates and missing values
- Removed special characters
- Lemmatization with WordNetLemmatizer()
- *Train-Test Split*
- *TF-IDF vectorizer*

Base Model: Random Forest Classifier

- Same parameters

Evaluation Metrics:

- Accuracy, Precision, Recall, F1-Score

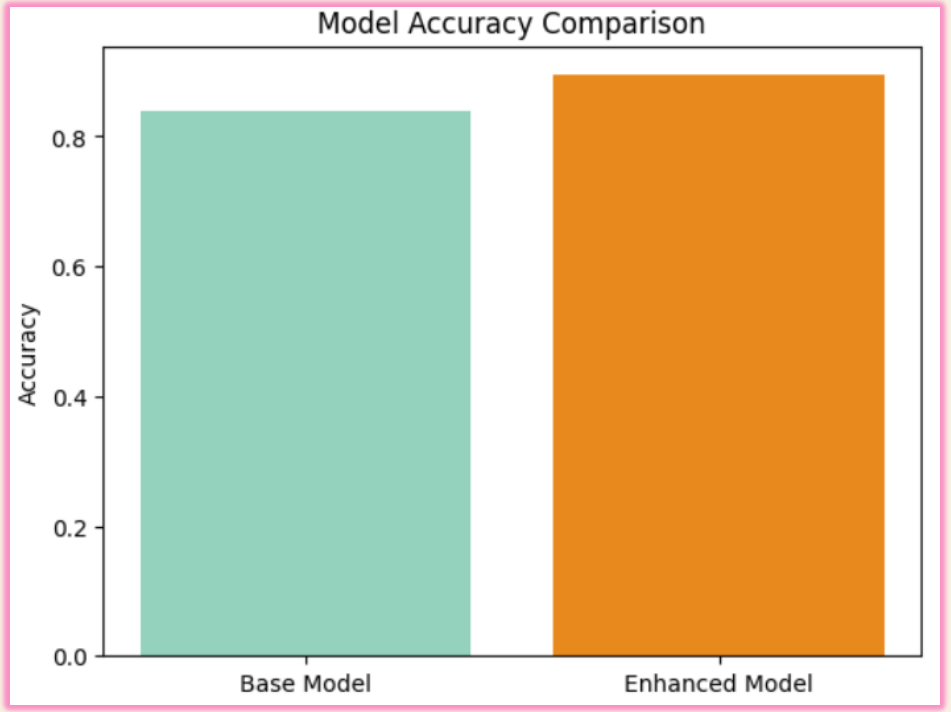
MODEL EVALUATION

Classification Report:

	precision	recall	f1-score	support
0	0.80	0.93	0.86	3529
1	0.91	0.74	0.82	3302
accuracy			0.84	6831
macro avg	0.85	0.84	0.84	6831
weighted avg	0.85	0.84	0.84	6831

Enhanced Model Classification Report:

	precision	recall	f1-score	support
0	0.89	0.90	0.90	3233
1	0.90	0.89	0.89	3209
accuracy			0.89	6442
macro avg	0.89	0.89	0.89	6442
weighted avg	0.89	0.89	0.89	6442



USING A TRANSFORMER-BASED MODEL

IMPLEMENTATION DETAILS

Data Preprocessing:

- *Removed duplicates and missing values*
- *Removed special characters*
- *Lemmatization with WordNetLemmatizer()*
- *Train-Test Split*
- Tokenization with Autotokenizer and a tokenize_function with padding, truncation and batch processing

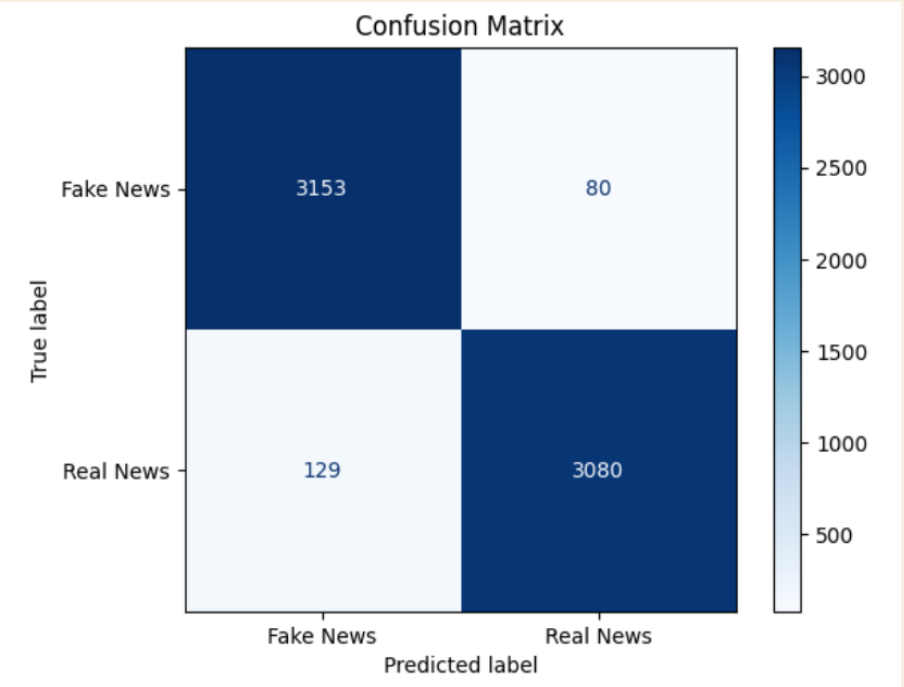
Model:

- DistilBERT (lightweight transformer model)
- Fine-tuned for binary classification
- Training and Evaluation with Hugging Faces' Trainer class

Evaluation Metrics:

- Accuracy, Precision, Recall, F1-Score

MODEL EVALUATION



Classification Report:

	precision	recall	f1-score	support
Fake News	0.96	0.98	0.97	3233
Real News	0.97	0.96	0.97	3209
accuracy			0.97	6442
macro avg	0.97	0.97	0.97	6442
weighted avg	0.97	0.97	0.97	6442

FINAL TIPS & TAKEAWAYS

DistilBERT Classifier with preprocessed data worked better

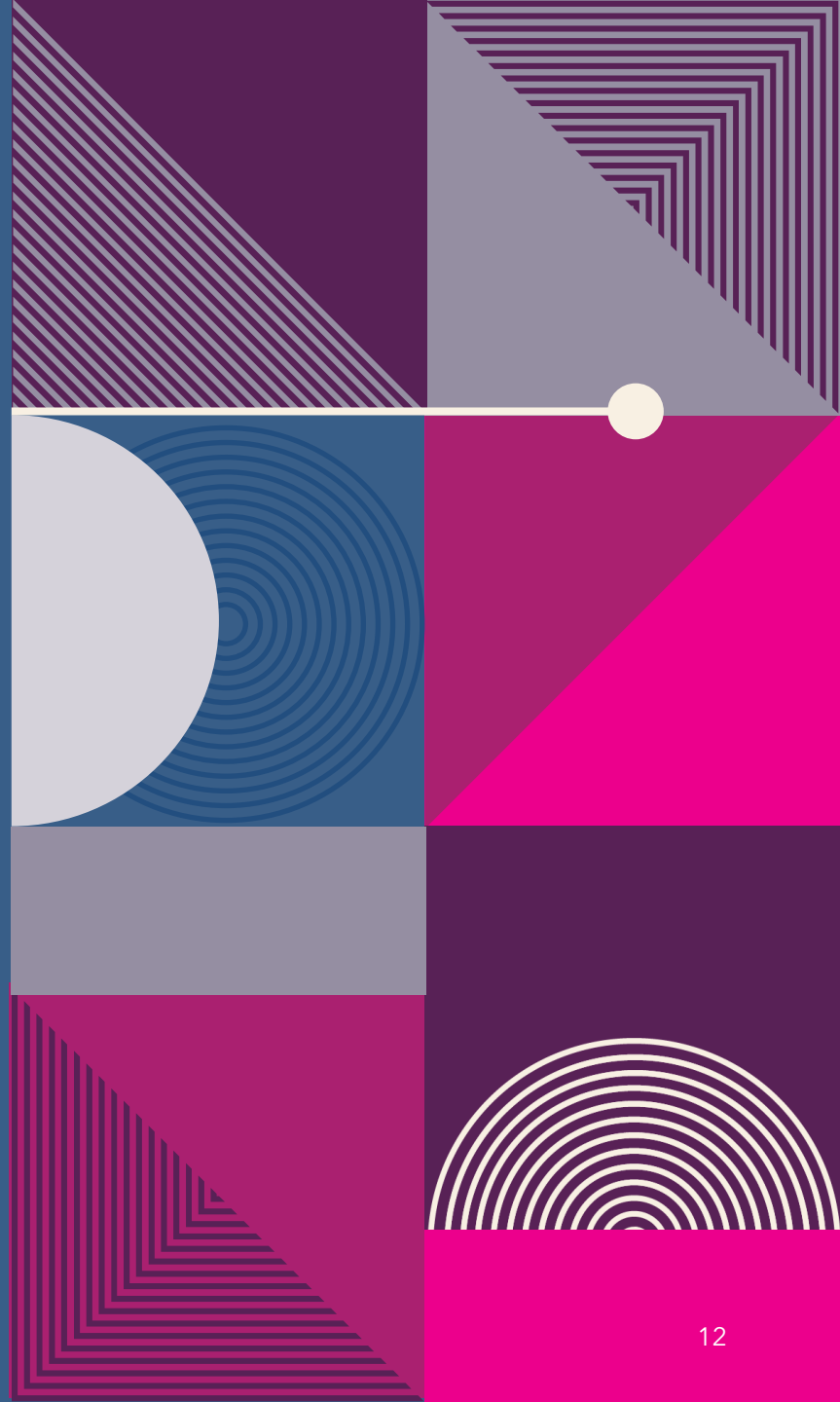
From 85% to 97% accuracy

Preprocessing is key! (again)

And making sure you work compatible packages

Next Steps:

Experiment with other token importance methods and transformer models



THANK YOU