

# FAKE NEWS DETECTION

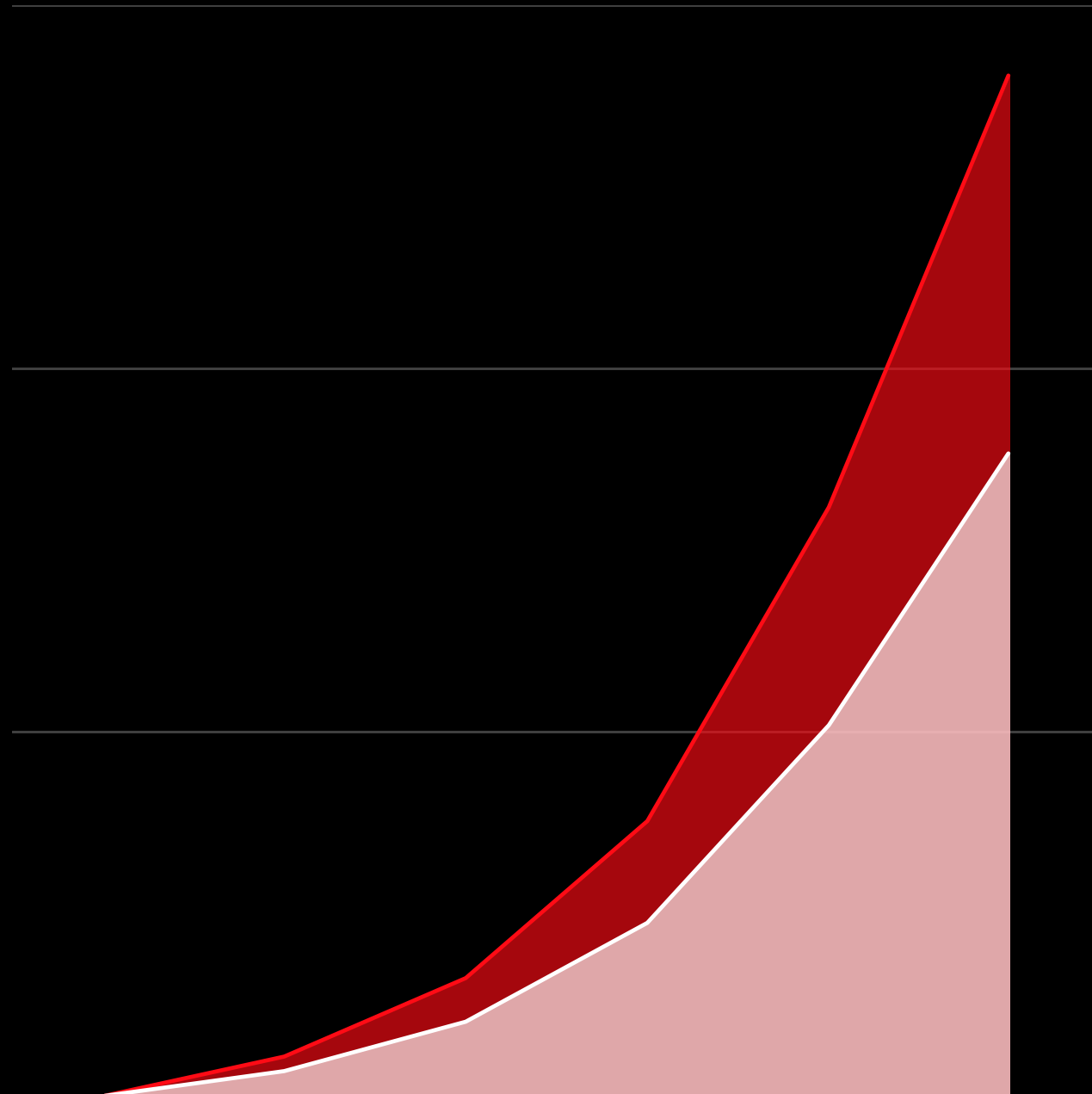
M. TALHA KAYA  
EYMEN GÜL



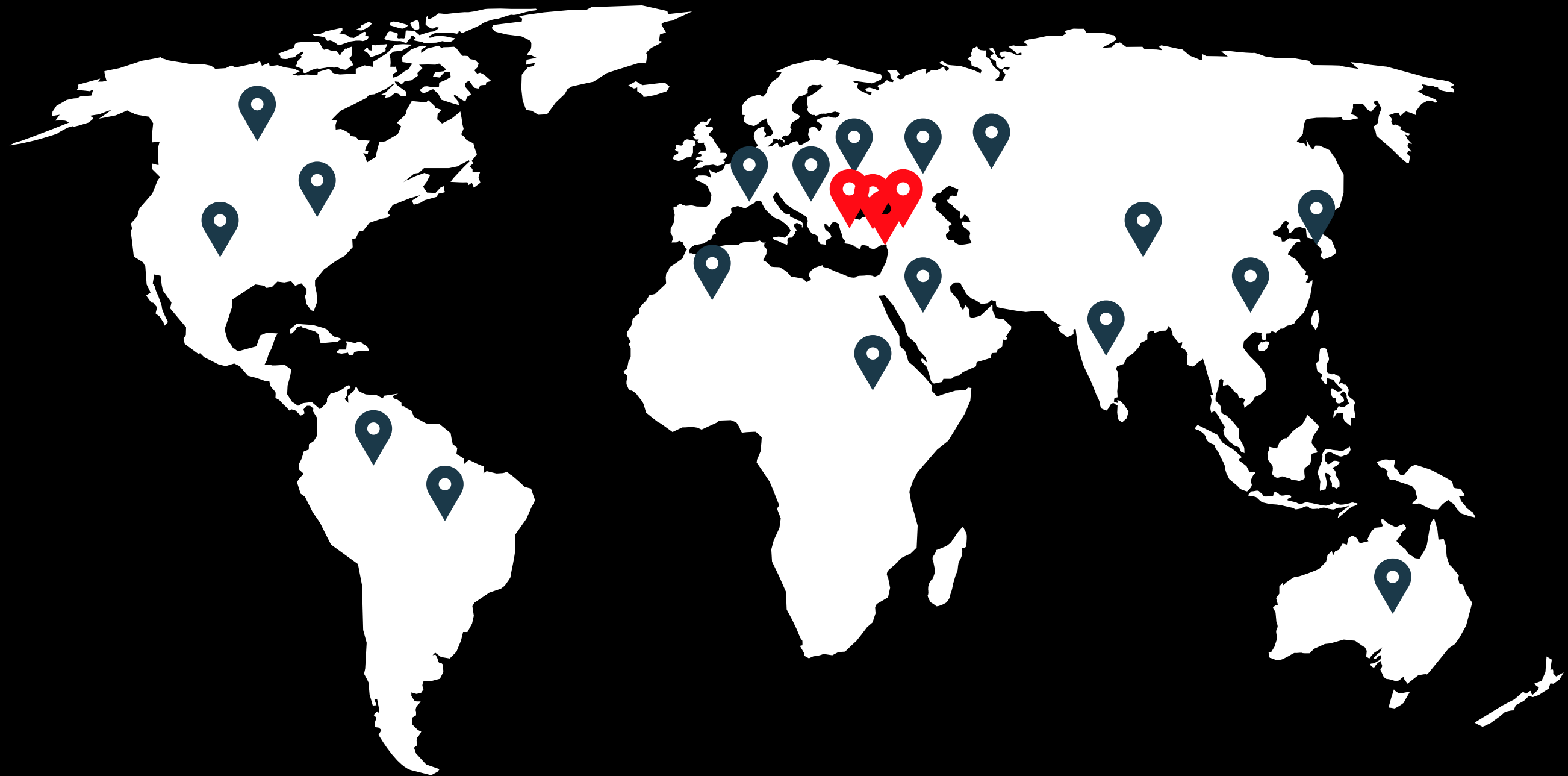
**FAKE NEWS**

Twitter and similar social media platforms significantly contribute to the spread of fake news. **Users' ability to quickly share unverified information** leads to the rapid dissemination of misleading tweets, often containing deceptive headlines, images, videos, or assumptions.

▪ REAL ▪ FAKE



**Fake news spreads  
70% faster than  
fact-based news on  
Facebook and Twitter  
(MIT, 2018).**



In 2019 alone, it was determined that fake news potentially influenced elections in at least **70 countries** worldwide (Oxford University, 2019).



# THE SPREAD OF FAKE NEWS DURING THE RECENT EARTHQUAKE IN TURKEY HAS LED TO A DECREASE IN ACCESS TO ACCURATE INFORMATION



THIS MUST BE PREVENTED.



# WE WORK WITH 2 DATASETS. FAKE ONE LABELED AS 0 TRUE ONE LABELED AS 1.

1	Donald Trump just couldn't wish all Americans ...	0
2	House Intelligence Committee Chairman Devin Nu...	0
3	On Friday, it was revealed that former Milwauk...	0
4	On Christmas day, Donald Trump announced that ...	0
5	Pope Francis used his annual Christmas Day mes...	0

21413	BRUSSELS (Reuters) - NATO allies on Tuesday we...	1
21414	LONDON (Reuters) - LexisNexis, a provider of l...	1
21415	MINSK (Reuters) - In the shadow of disused Sov...	1
21416	MOSCOW (Reuters) - Vatican Secretary of State ...	1
21417	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	1

44898 news used  
to create a model.  
There is no null  
values in  
datasets.

We completely excluded  
10 pieces of data from this  
work. To try when the  
model is complete.



# DATA CLEANING IS DONE BY WORDOPT FUNCTION

```
def wordopt(text):
    text = text.lower()
    text = re.sub('\[.*?\]', "", text)
    text = re.sub("\\W", " ", text)
    text = re.sub(
        'https?://\S+|www\.\S+', "", text)
    text = re.sub('<.*?>+', "", text)
    text = re.sub('[%s]' %
        re.escape(string.punctuation), "",
        text)
    text = re.sub('\n', "", text)
    text = re.sub('\w*\d\w*', "", text)
    return text
```

# Feature Extraction

We utilized the TF-IDF (Term Frequency-Inverse Document Frequency) method for feature extraction.

It's a measure used to evaluate the importance of a word in a document or a collection of texts.

**Term Frequency (TF)** measures the usage frequency of a certain word in a specific document.

It's usually calculated as the number of times a word appears in the document divided by the total number of words in that document.

**Inverse Document Frequency (IDF)** measures the inverse prevalence of a word in the collection of documents.

Words that are commonly used tend to contain less information, so IDF helps to reduce the weight of these words.

# MODEL TRAINING, EVALUATION, AND VALIDATION

## Logistic Regression

	Precision	Recall	f1-score	Support
0	0.99	0.98	0.99	5869
1	0.98	0.99	0.99	5353
accuracy			0.99	11222
macro avg.	0.99	0.99	0.99	11222
weighted avg.	0.99	0.99	0.99	11222

**Precision** : Precision: It's the ratio of correctly predicted positive observations to the total predicted positives.

**Recall (Sensitivity)**: It's the ratio of correctly predicted positive observations to all the actual positives.

**F1-Score**: It's the harmonic mean of Precision and Recall and gives a better measure of the incorrectly classified cases than the Accuracy Metric.

**Support**: This represents the actual number of occurrences of the class in the specified dataset.

# MODEL TRAINING, EVALUATION, AND VALIDATION

Decision Tree Classification

	Precision	Recall	f1-score	Support
0	0.99	0.99	0.99	5869
1	0.99	0.99	0.99	5353
accuracy			0.99	11222
macro avg.	0.99	0.99	0.99	11222
weighted avg.	0.99	0.99	0.99	11222

Random Forest Classifier

	Precision	Recall	f1-score	Support
0	0.99	0.99	0.99	5869
1	0.99	0.99	0.99	5353
accuracy			0.99	11222
macro avg.	0.99	0.99	0.99	11222
weighted avg.	0.99	0.99	0.99	11222



## FAKE OR REAL ?

Bir metin girin...

Gönder

Our model's performance is excellent. We've implemented it on a local site for operation.

**THANK YOU**

**MAT-388E**

**FAKE&REAL NEWS**

**DETECTION**

**EYMEN GUL-090190746**

**M. TALHA KAYA-90160339**



**LET'S  
EXPLORE  
THE  
WEBSITE**

