WHO Said That?

AI-Based Covid-19 Fake News Detection System

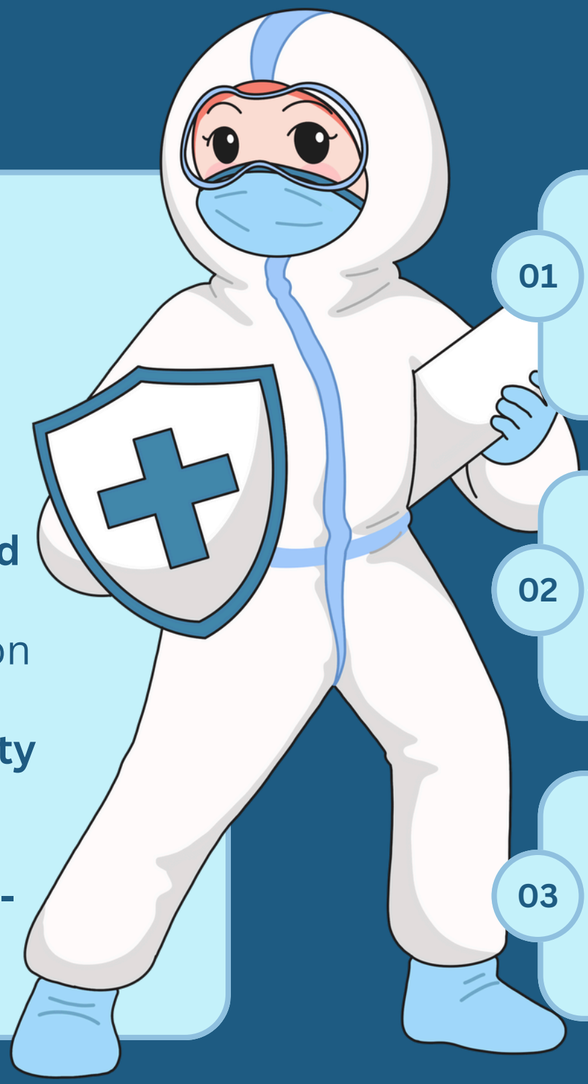
- Syed M. Ali Raza Naqvi (01-134222-149)
- Qurat-ul-Ain Fatima (01-134222-121)

Introduction

- Covid-19 has caused a global health crisis, but an “**infodemic**” of fake news has worsened the situation.
- Misinformation about COVID-19 leads to fear, confusion, and risky behaviors.
- Traditional manual verification is slow and not scalable.
- We need an automated, AI-based solution to detect and flag fake news instantly.

Project Objectives

Our project aims to **safeguard public health** through **AI-powered solutions** focused on **fake news detection**, **early identification**, and **community awareness**.
The goal is to build a **smart system** that classifies **COVID-19 news** as “**Real**” or “**Fake**.”



01 Build an **automated tool** using **NLP** and **Machine Learning** to detect **misleading COVID-19 news.**

02 Design a **simple web interface** for users to **verify news** in real time with accurate results.

03 Help fight **misinformation** and support **public awareness** during **health crises**.

Project Outcomes

Our project successfully delivered a working prototype that **combines AI and NLP** to combat misinformation during global health emergencies. The outcomes reflect the effectiveness, usability, and practical value of the solution.

01 **Developed a high-accuracy fake news detection model** using Naive Bayes and TF-IDF, achieving 91.09% accuracy in classifying COVID-19 news articles.

02 **Built an interactive web application** where users can input tweets and instantly receive a “Real” or “Fake” classification.

03 **Enhanced public awareness** by providing a tool that promotes critical thinking and helps users identify potentially harmful misinformation.

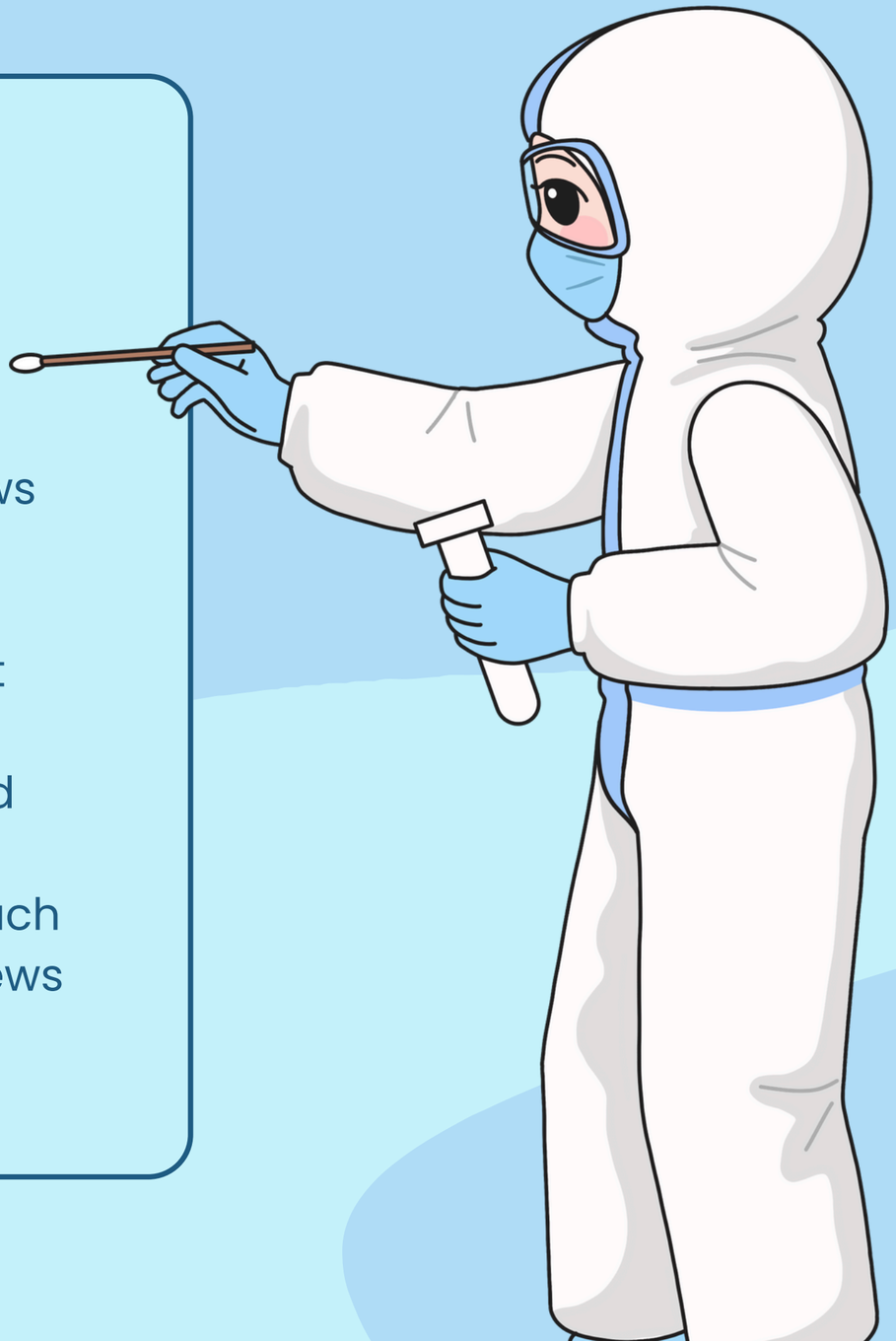
04 **Benchmarking and testing validated the model’s robustness**, demonstrating reliable performance across multiple evaluation metrics (precision, recall, F1-score).

05 **Deployed a modular AI system** that can be extended to detect misinformation in other domains such as politics, climate change, and health.

Dataset Overview

Source: Github (English Language Covid-19 Tweets)
Dataset: Constraint COVID-19 Fake News Dataset
Fields in the Dataset:

1. ***id:*** A unique identifier assigned to each tweet or news post. It helps keep track of every individual record without repeating or confusing entries.
2. ***tweet:*** The actual text content of the tweet or post. It contains COVID-19 related information, statements, claims, opinions, or reports which need to be verified for authenticity.
3. ***label:*** The ground truth classification assigned to each tweet. It tells whether the information is real (true news based on reliable sources) or fake (misinformation, rumors, or false claims).



Dataset Details

01

Classes:

- **Real** - 5,580 tweets
- **Fake** - 5,089 tweets

Total Records: 10,669 tweets

02

Train-Test Split:

- **Training Split** - 80%
- **Testing Split** - 20%

The split was randomized to ensure unbiased sampling and generalization.

03

Preprocessing Steps:

- Lowercased text to standardize and reduce dimensions.
- Removed URLs, punctuation, and stopwords to clean noise.
- Tokenized text into processable units.
- Applied stemming/lemmatization to unify word forms.
- Used TF-IDF to convert text into numerical features for modeling.

Model Selection

Model Benchmarking

	Model	Accuracy	Precision	Recall	F1	Train_s	Predict_s
0	Linear SVM	0.924455	0.924701	0.924455	0.924483	0.412512	0.051914
1	LogReg	0.910436	0.911463	0.910436	0.910485	2.641314	0.147289
2	Naive Bayes	0.904984	0.905427	0.904984	0.905028	0.591475	0.100052
3	RandomForest	0.896417	0.898013	0.896417	0.896471	20.416697	0.324102

We selected **Naive Bayes**. But Why?

Why we Chose Naive Bayes?

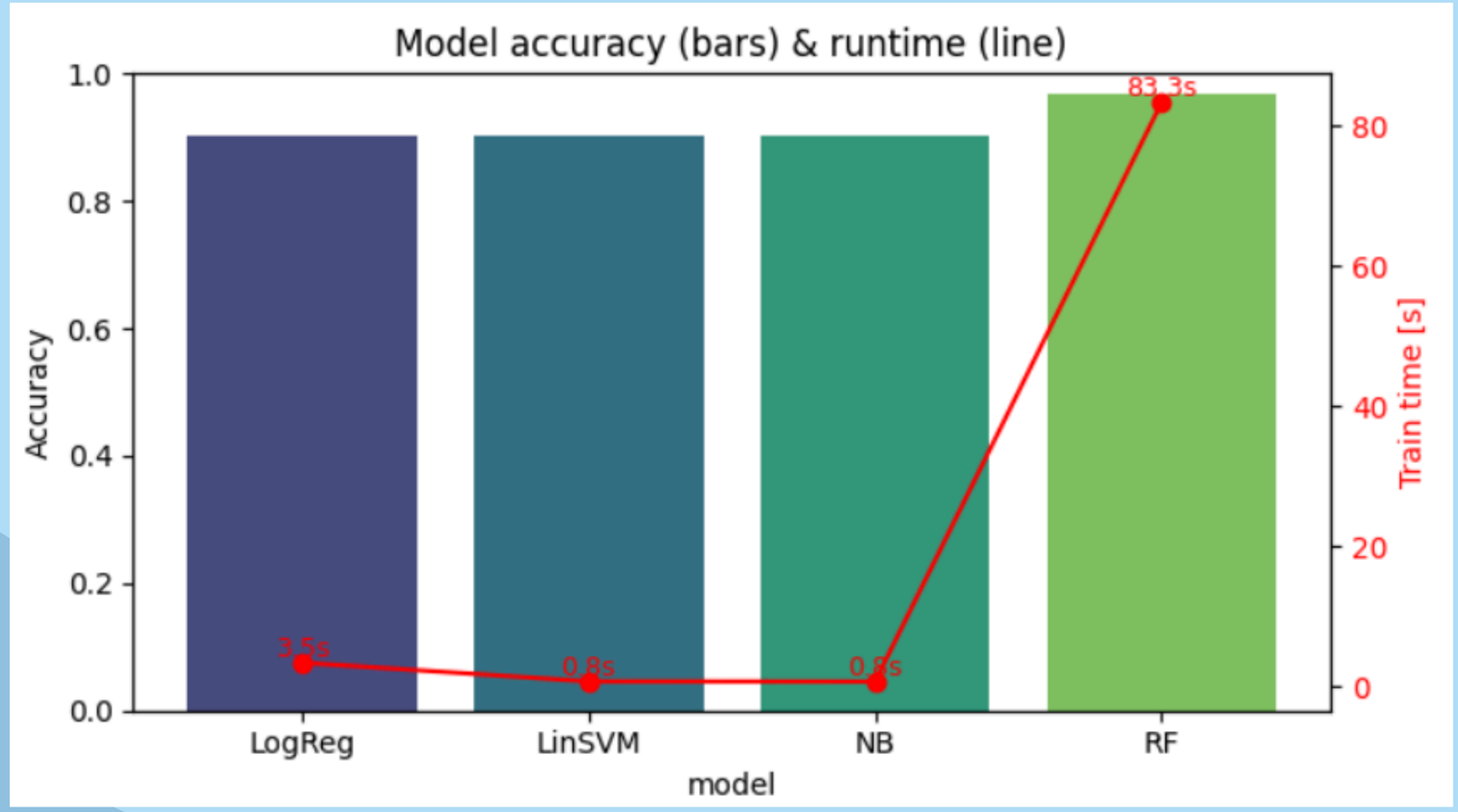
Real-Time Accuracy Matters
SVM and Logistic Regression performed better on paper but gave inconsistent results in live testing.

Naive Bayes: Reliable & Efficient

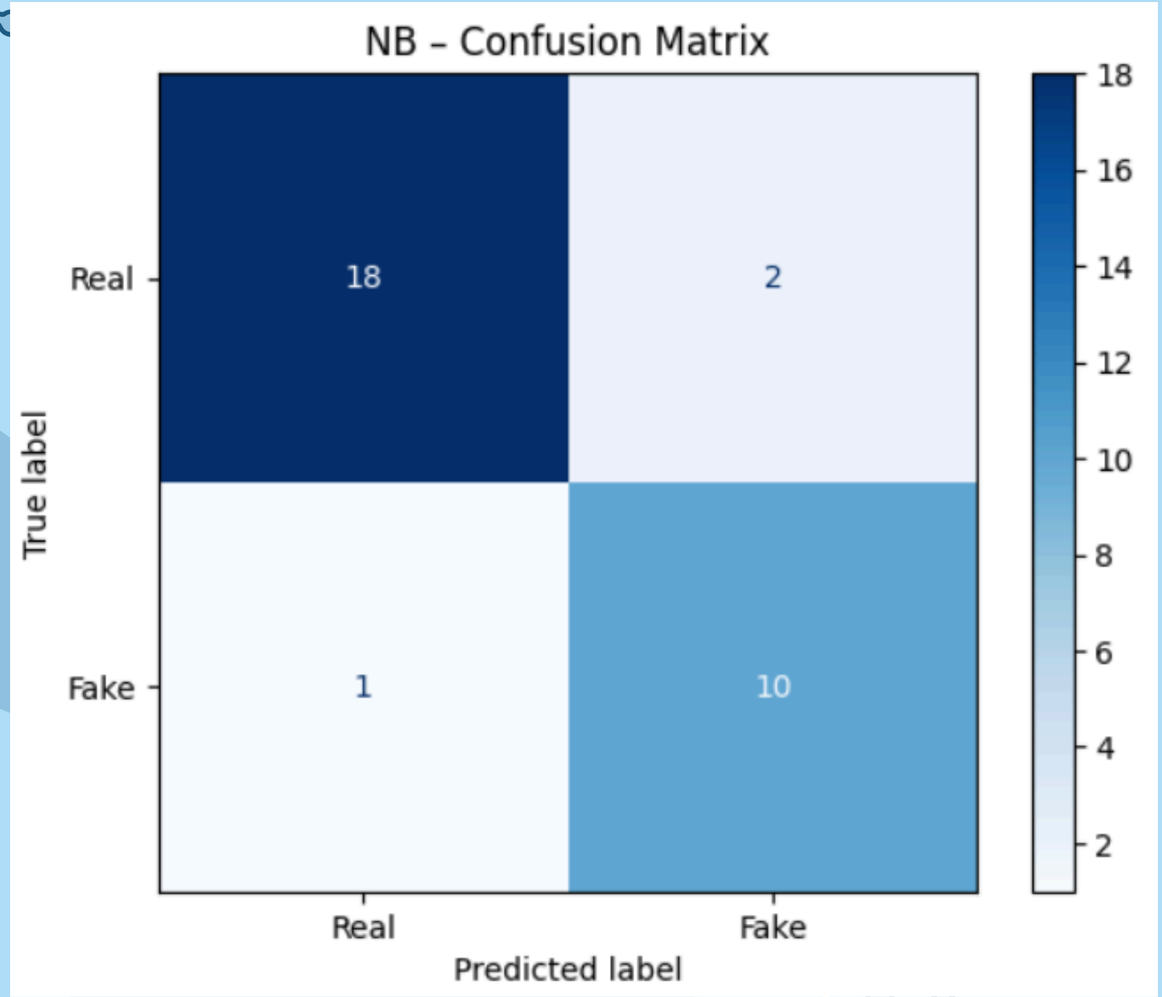
- Consistent predictions during runtime
- High accuracy and F1-score
- Fast training and prediction times
- Performs well on short, noisy text (e.g., tweets)

Final Choice: Naive Bayes
A balance of performance, speed, and real-world reliability.

Why we Chose Naive Bayes?



Confusion Matrix




```
=== NB metrics ===
              precision    recall  f1-score   support

        fake      0.833     0.909     0.870        11
        real      0.947     0.900     0.923        20

    accuracy                          0.903        31
   macro avg      0.890     0.905     0.896        31
weighted avg      0.907     0.903     0.904        31
```

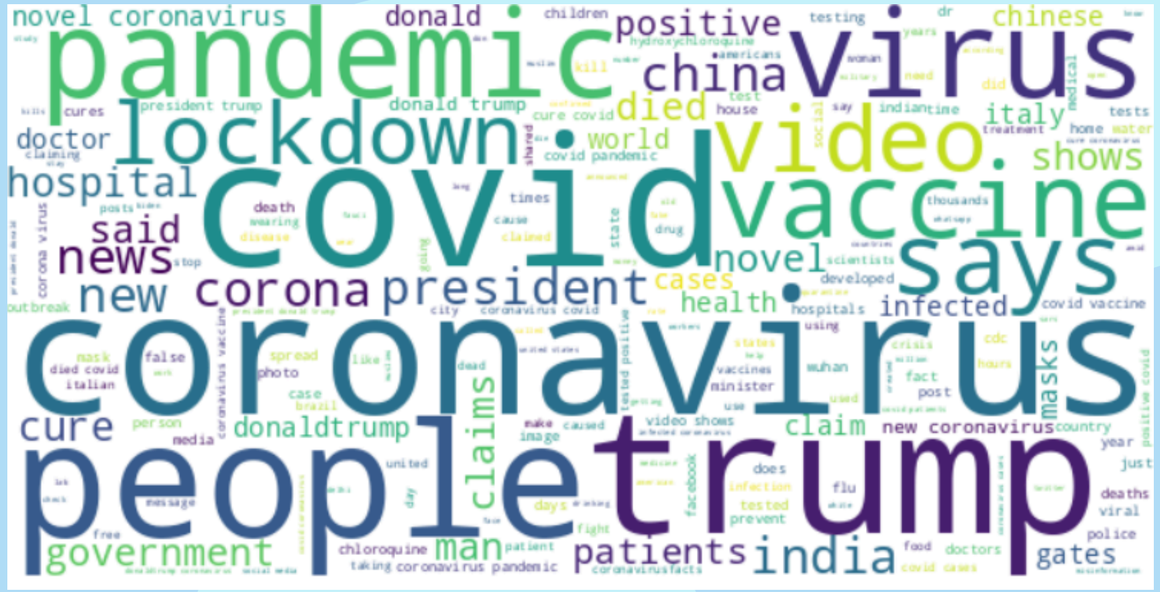
TF-IDF + Word Cloud

Why we used TF-IDF?

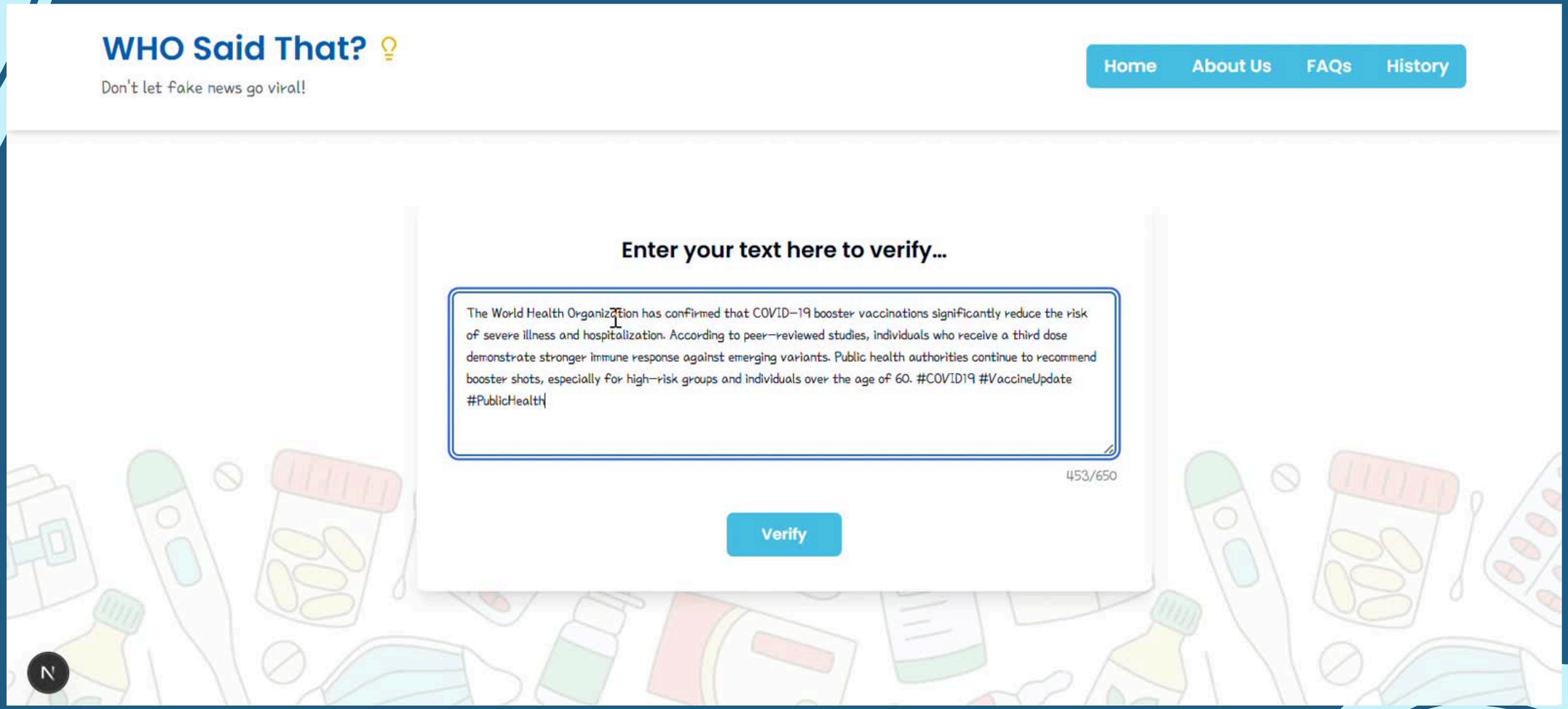
- Converts tweet text into numeric features for ML models.
- Emphasizes important words, downplays common ones.
- Works well with short, noisy text like tweets.
- Supports n-grams (e.g., "covid hoax").
- Fast and lightweight, ideal for real-time use.

What is TF-IDF?

- TF-IDF stands for Term Frequency–Inverse Document Frequency.
- It’s a numerical representation of text that reflects how important a word is to a document in a collection.



WHO Said That? – Live Demo



If this video doesn't play, watch the demo video on Linkedin here

THANK YOU!