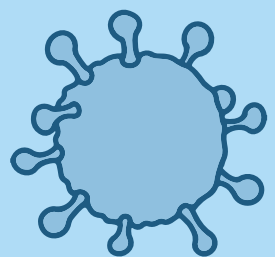


# WHO Said That?

## 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.





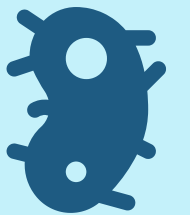
# Project Objectives

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**.”

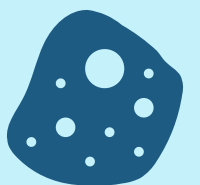
01

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



02

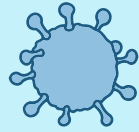
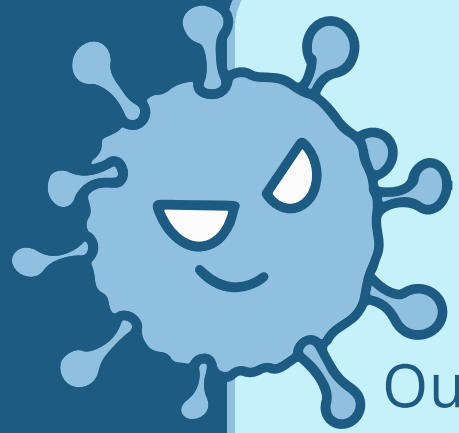
Design a **simple web interface** for users to **verify news** in real time with accurate results.



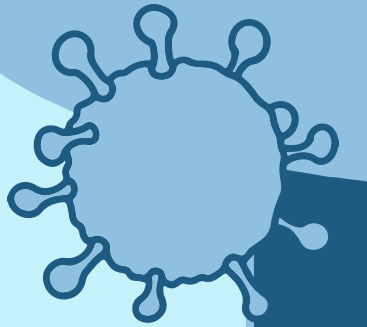
03

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

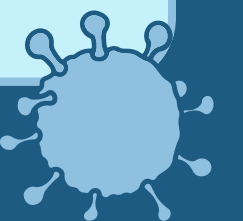
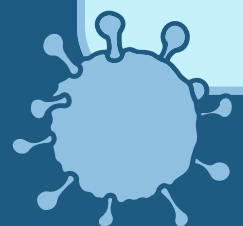
**Enhanced public awareness** by providing a tool that promotes critical thinking and helps users identify potentially harmful misinformation.

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.





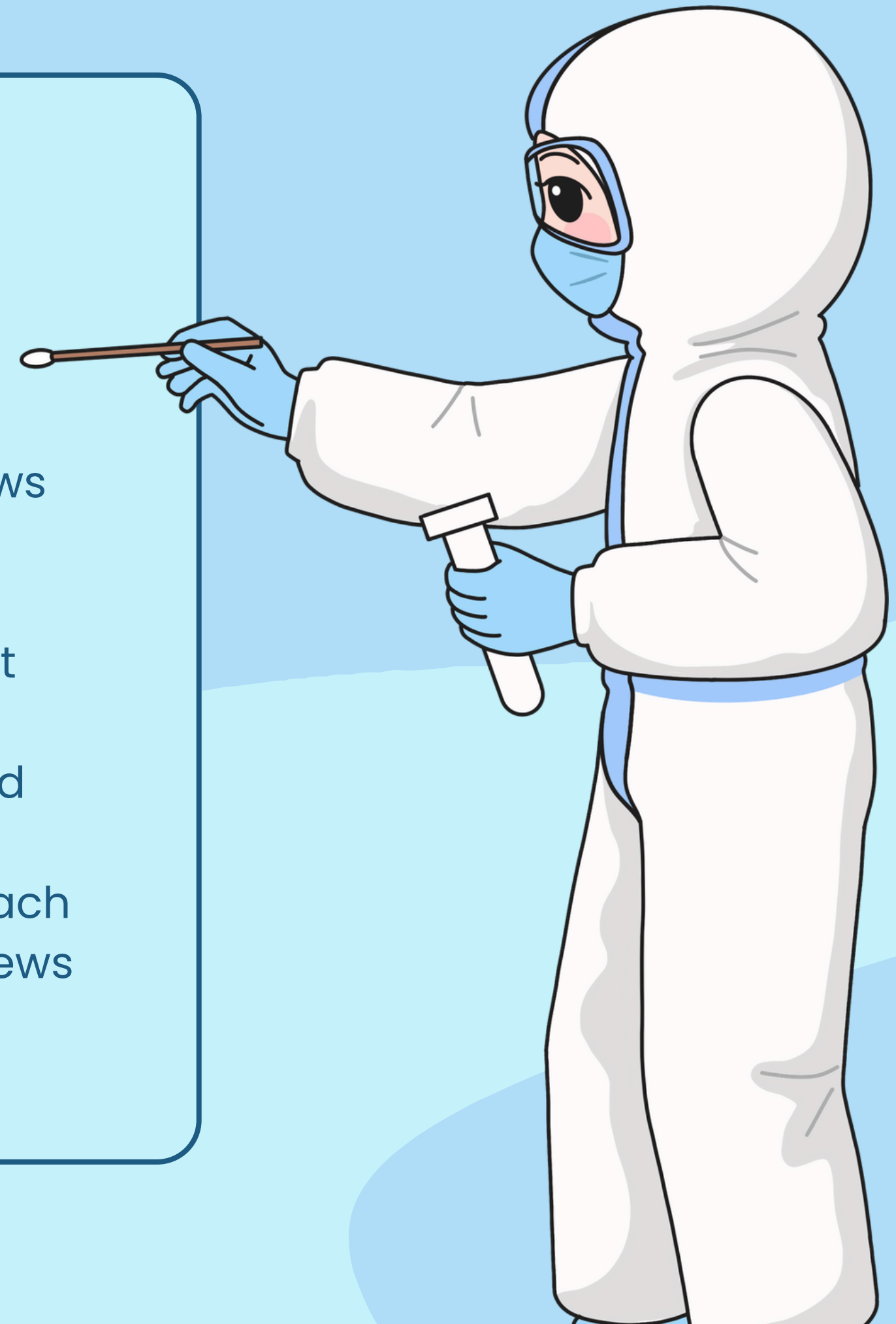
# Dataset Overview

**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

	acc	prec	rec	f1	fit_s	pred_ms
model						
LogReg	0.870968	0.882272	0.870968	0.872949	3.094834	4.049674
LinSVM	0.903226	0.906904	0.903226	0.904089	1.013617	3.961459
NB	0.870968	0.882272	0.870968	0.872949	1.126694	3.056315
RF	0.903226	0.923963	0.903226	0.905109	60.623400	121.636780

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**

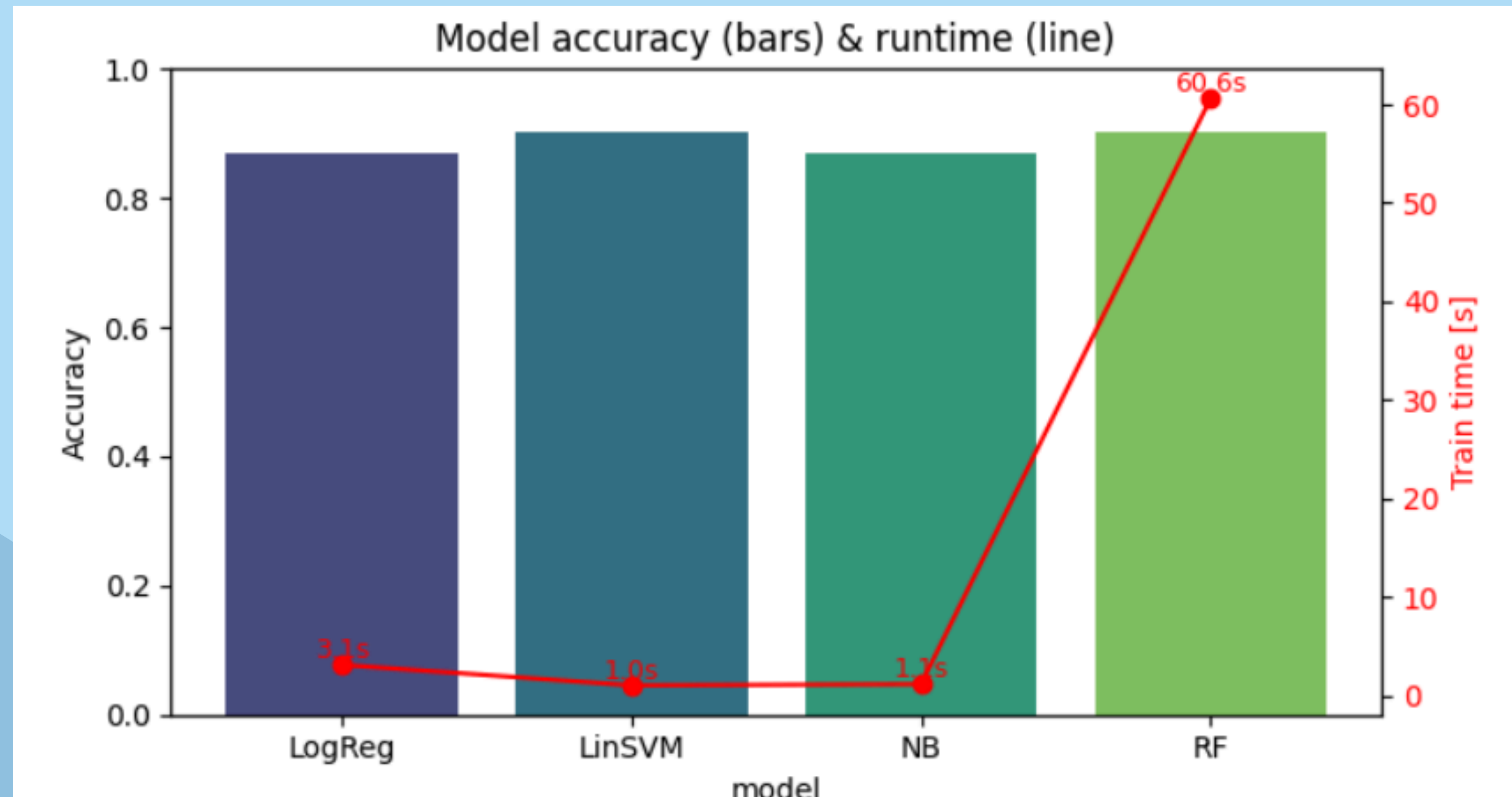
- Consistent predictions during runtime
- High accuracy and F1-score
- Fast training and prediction times
- Performs well on short, noisy text (e.g., tweets)

## **Final Choice: Naive Bayes**

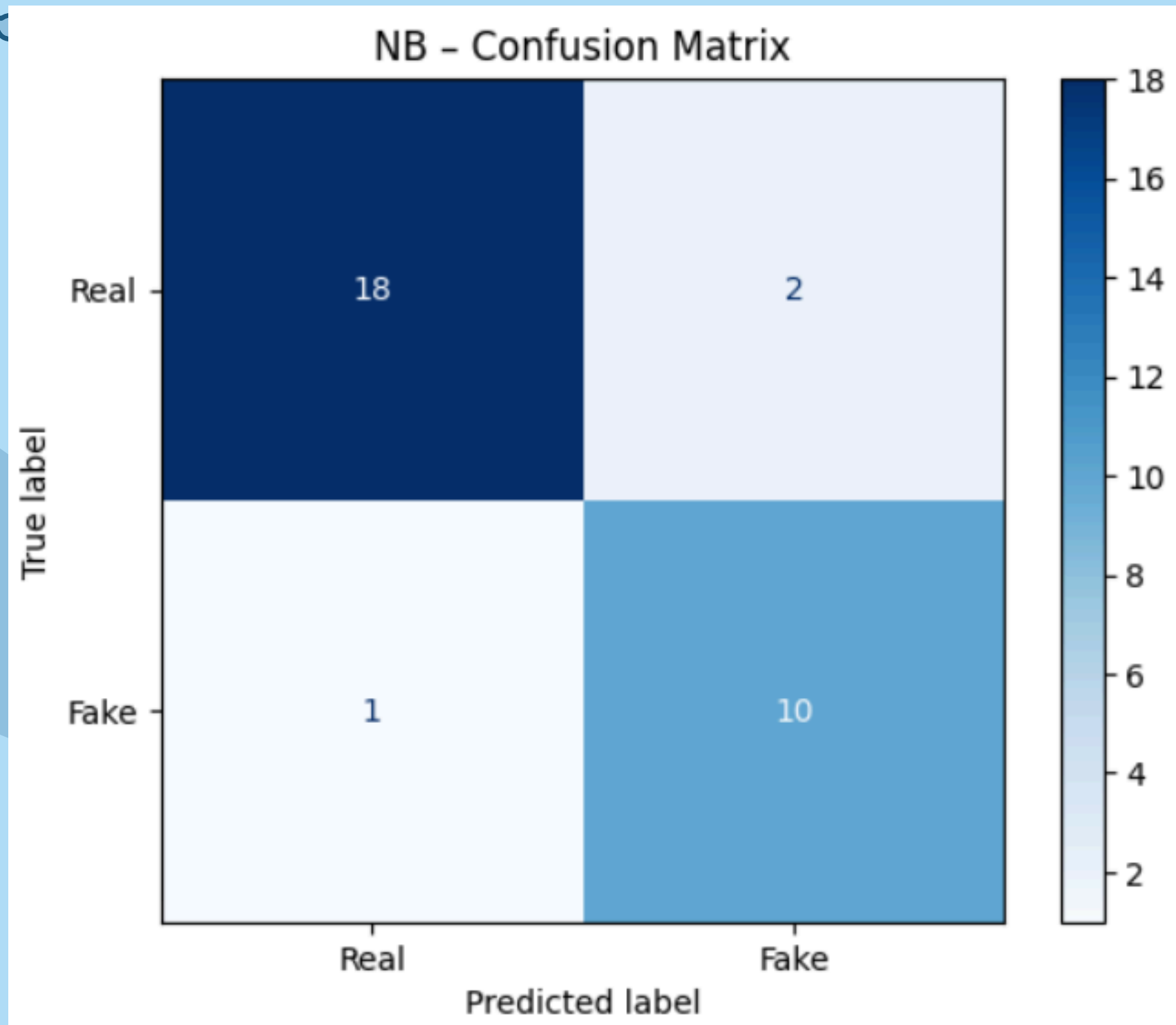
A balance of performance, speed, and real-world reliability.



# Why we Chose Naive Bayes?



# Confusion Matrix



=== NB Test Performance ===

	precision	recall	f1-score	support
fake	0.769	0.909	0.833	11
real	0.944	0.850	0.895	20
accuracy			0.871	31
macro avg	0.857	0.880	0.864	31
weighted avg	0.882	0.871	0.873	31

## TF-IDF + Word Cloud

## What is TF-IDF?

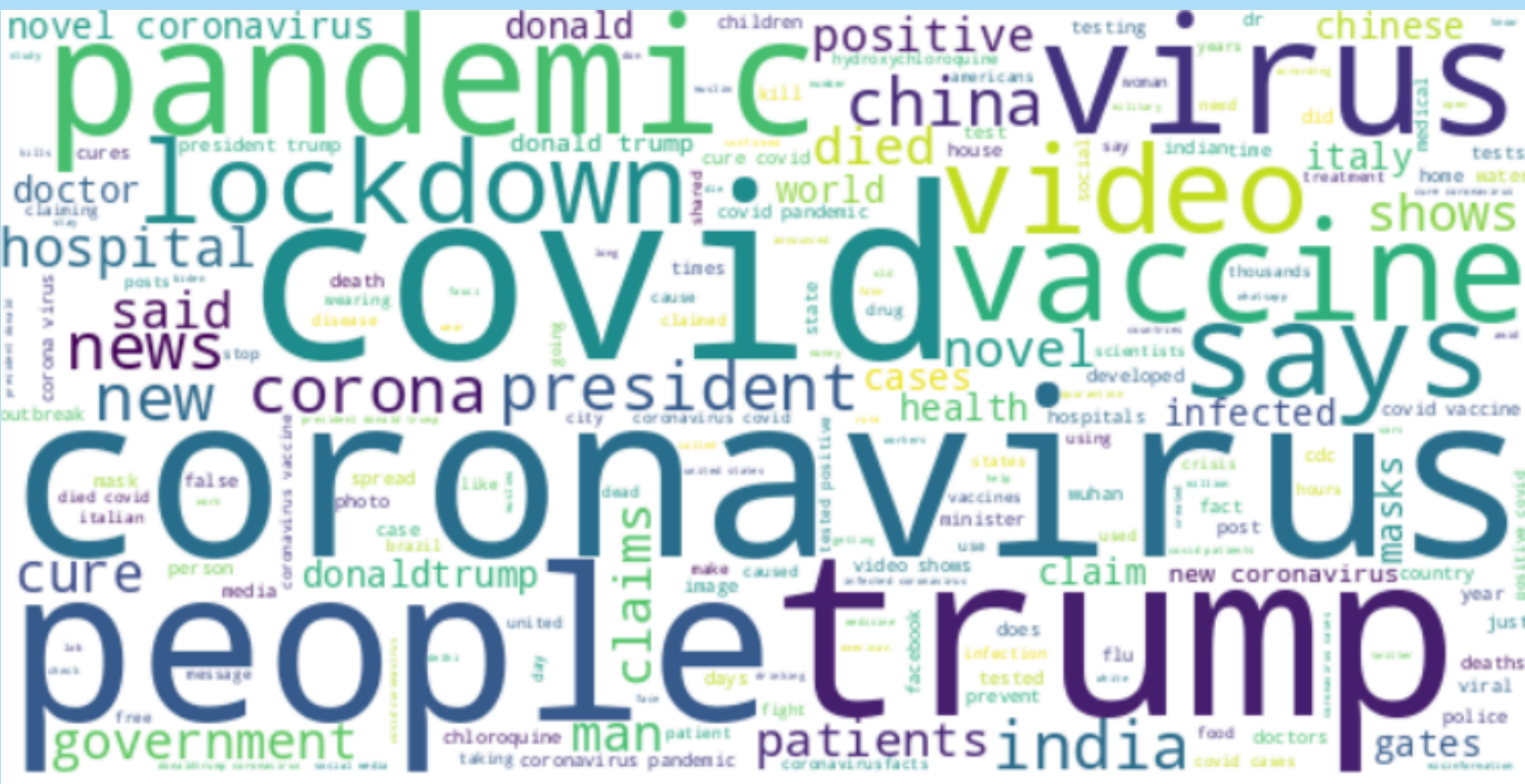
- TF-IDF stands for Term Frequency–Inverse Document Frequency.
- It's a numerical representation of text that reflects how important a word is to a document in a collection.

- ## 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.

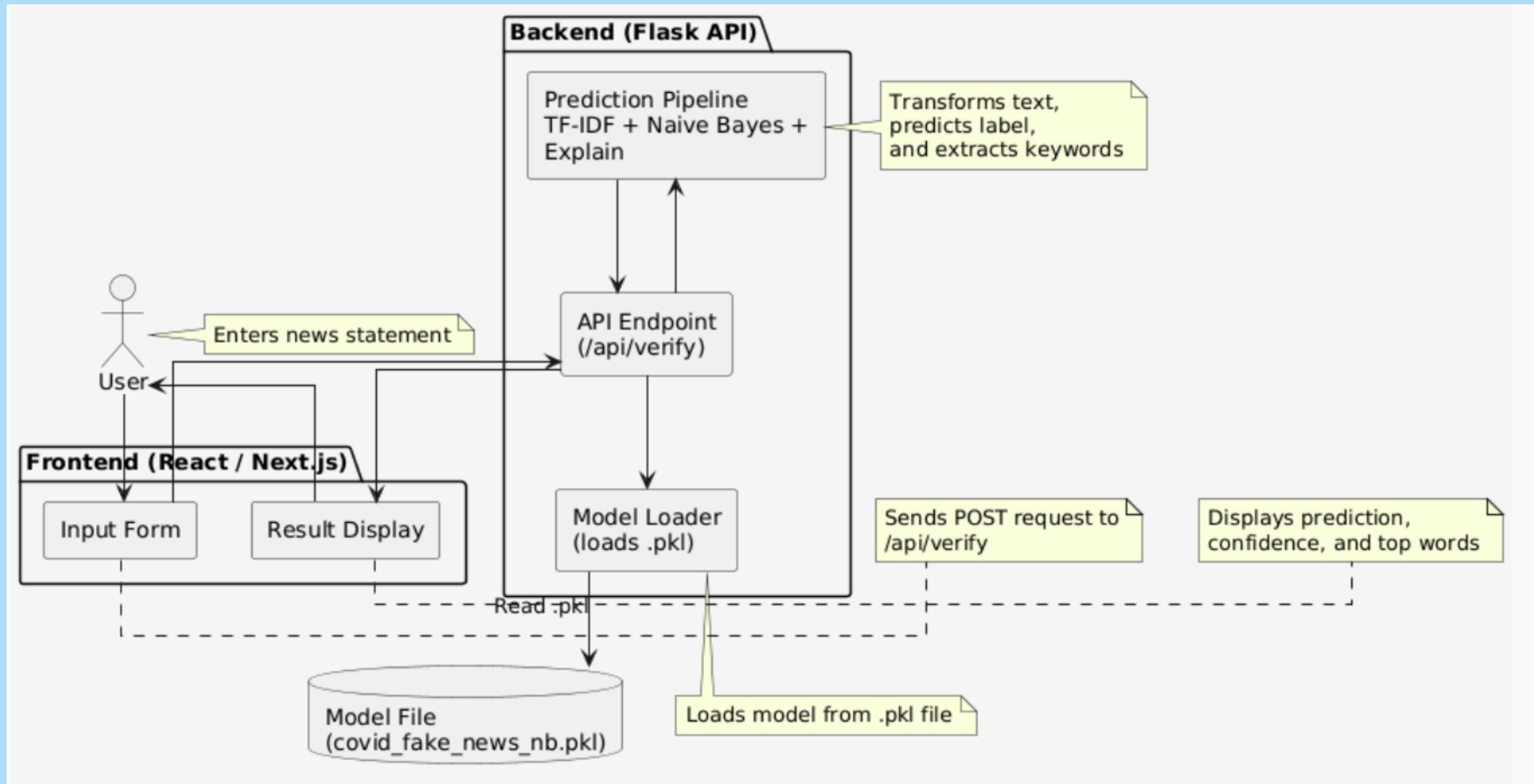
## Why we used TF-IDF?

- 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.

- ## 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.



# System Architecture



# WHO Said That? – Live Demo

WHO Said That? 💡

Don't let fake news go viral!

Home

About Us

FAQs

History

Enter your text here to verify...

The World Health Organization has confirmed that COVID-19 booster vaccinations significantly reduce the risk of severe illness and hospitalization. According to peer-reviewed studies, individuals who receive a third dose demonstrate stronger immune response against emerging variants. Public health authorities continue to recommend booster shots, especially for high-risk groups and individuals over the age of 60. #COVID19 #VaccineUpdate #PublicHealth

453/650

Verify

If this video doesn't play, watch the demo video on LinkedIn [here](#)





**THANK YOU!**