

CAPSTONE PROJECT

PHASE #5

# THE PROJECT

## Business Problem

- 70% of Americans say they came across 'fake news' every day on social media in 2023 (source: D.Georgiev, 2024)
- How can we know if an article is a fake or a real news?

## Data

- 45,000 news classified as "Fake" (label 0) or "Real (label 1)
- Accessible publicly from Kaggle: <https://www.kaggle.com/datasets/aadyasingh55/fake-news-classification/data>

## Stakeholder

- Head of Innovation of Social Media companies -> Meta, Snapchat, TikTok, X (formerly Twitter)

## Solution

- A trusted and accurate Machine Learning model that can identify a fake and a real news

# TRAINED MODELS

#1 - Multinomial Naive  
Bayes with Count  
Vectorizer



#1.1 - with  $\alpha = 0.001$

#1.2 - with  $\alpha = 0.01$

#1.3 - with  $\alpha = 0.1$

#2 - Multinomial Naive  
Bayes with TF-IDF  
Vectorizer



#2.1 - with  $\alpha = 0.001$

#2.2 - with  $\alpha = 0.01$

#2.3 - with  $\alpha = 0.1$

# FINAL MODEL

**Multinomial Naive  
Bayes with Count  
Vectorizer & alpha =  
0.001**

## Why?

- **Best overall performance across key metrics (Accuracy, Cohen's Kappa, Matthew's Correlation Coefficient)**
- **Balanced precision and recall without sacrificing consistency**
- **Robust stability across alpha variations ensures reliable predictions**

# DEPLOYMENT

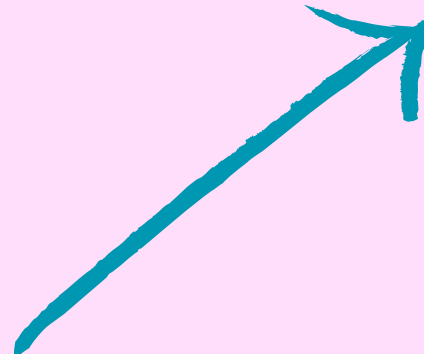
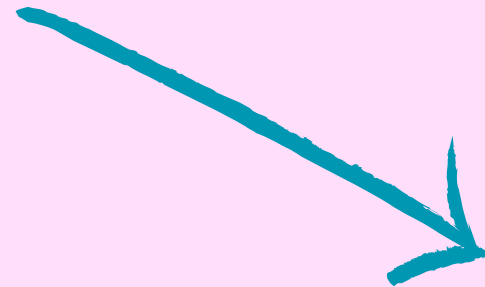
**Streamlit  
Library for  
front-end**

**New Input from  
the User**



**Cleaning,  
Preprocessing  
& Tokenization**

**ML Final Model  
with Pickle file**



# DEMO TIME!

*Link -> <https://youtu.be/CPYNyuilGCQ>.*

THANKS!

Q/A TIME!