PROJECT 3

FAKE NEWS DETECTION IN REDDIT NEWS DATABASE

CHALERMCHON WONGSOPA

WARINTORN NAWONG

AGENDA

- *Introduction : Who are we?*
- *Model Development Plan*
- *Research & Data Exploring*
- *Scenario Analysis*
- *Model Fine-tuning*
- *Performance & Reliability Test*
- *Summary*
- *Way Forward*

Who are we?

ADVANCE PUBLICATIONS

reddit | Charter COMMUNICATIONS | WARNER BROS. DISCOVERY | ACBJ

REDDIT IPO READINESS PLAN 2022

IPO Requirement List :
Fake News Reduction

IPO : Initial Public Offering.

Why does 'Fake News' matters?

26.0 %

SHARE OF AMERICAN ***VERY CONFIDENT*** IN THEIR ABILITY TO RECOGNIZE FAKE.

67.0 %

AMERICANS WHO **BELIEVE FAKE NEWS** CAUSES A GREAT DEAL OF CONFUSION.

38.2 %

AMERICANS WHO ***ACCIDENTALLY SHARED*** FAKE NEWS.

Reference : https://www.statista.com/topics/3251/fake-news/#topicHeader__wrapper

Which one is Fake News?

“Kim Yo Jong Confirms Kim Jong Un Is Dead.”

“North Korea leader Kim Jong-un 'suffered fever’.“

Reference : https://www.bbc.com/news/world-asia-62501152

Which one is Fake News?

A

“Kim Yo Jong Confirms Kim Jong Un Is Dead.”

B

“North Korea leader Kim Jong-un 'suffered fever.”

Reference : https://www.bbc.com/news/world-asia-62501152

Which one is Fake News?

“ Prolonged Use of Face Mask Causes Hypoxia. ”

“ Lincoln Memorial Defaced By BLM Protestors. “

Reference : https://www.techarp.com/internet/lincoln-memorial-blm-defaced/

Which one is Fake News?

“

Prolonged Use of Face Mask Causes Hypoxia.

”

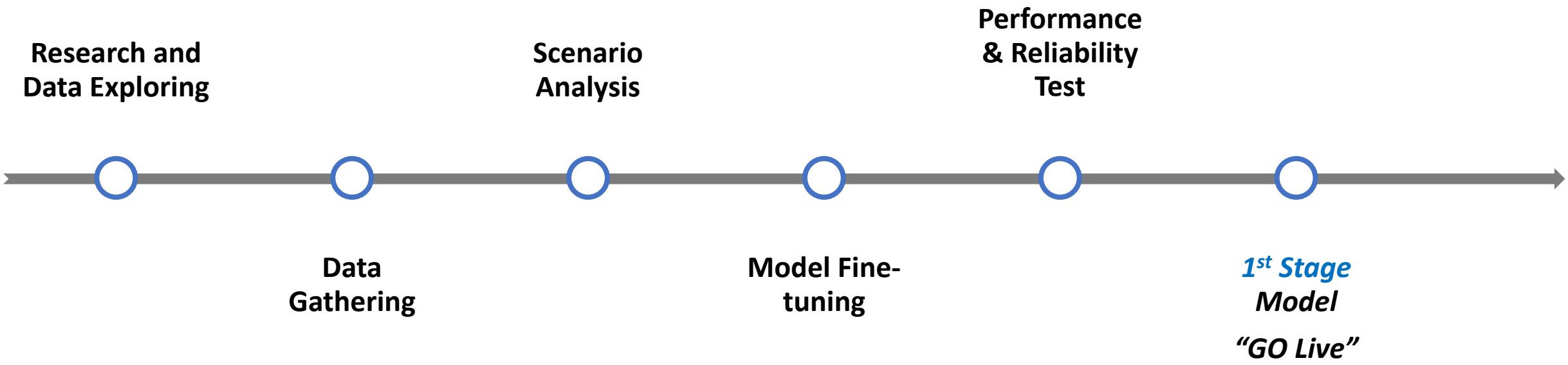
“

Lincoln Memorial Defaced By BLM Protestors.

“

Reference : https://www.techarp.com/internet/lincoln-memorial-blm-defaced/

Model Development Plan



Research & Data Exploring

subreddit

FAKE NEWS DATASET	*REAL NEWS DATASET*
r/fakenews	r/news
r/AteTheOnion	r/worldnews
r/TheOnion	r/UpliftingNews
	r/Coronavirus

Limitations of Using these data – HIGH BIAS

Difference In Timeline

Fake News

Real News

Difference In Topics

Fake News

Real News

Research & Data Exploring (Cont.)

FAKE NEWS DATASET

FILTER

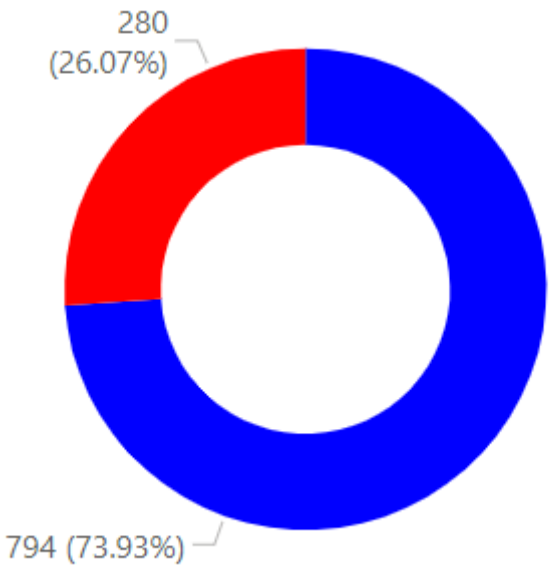
SAME TOPICS

REAL NEWS DATASET

FILTER

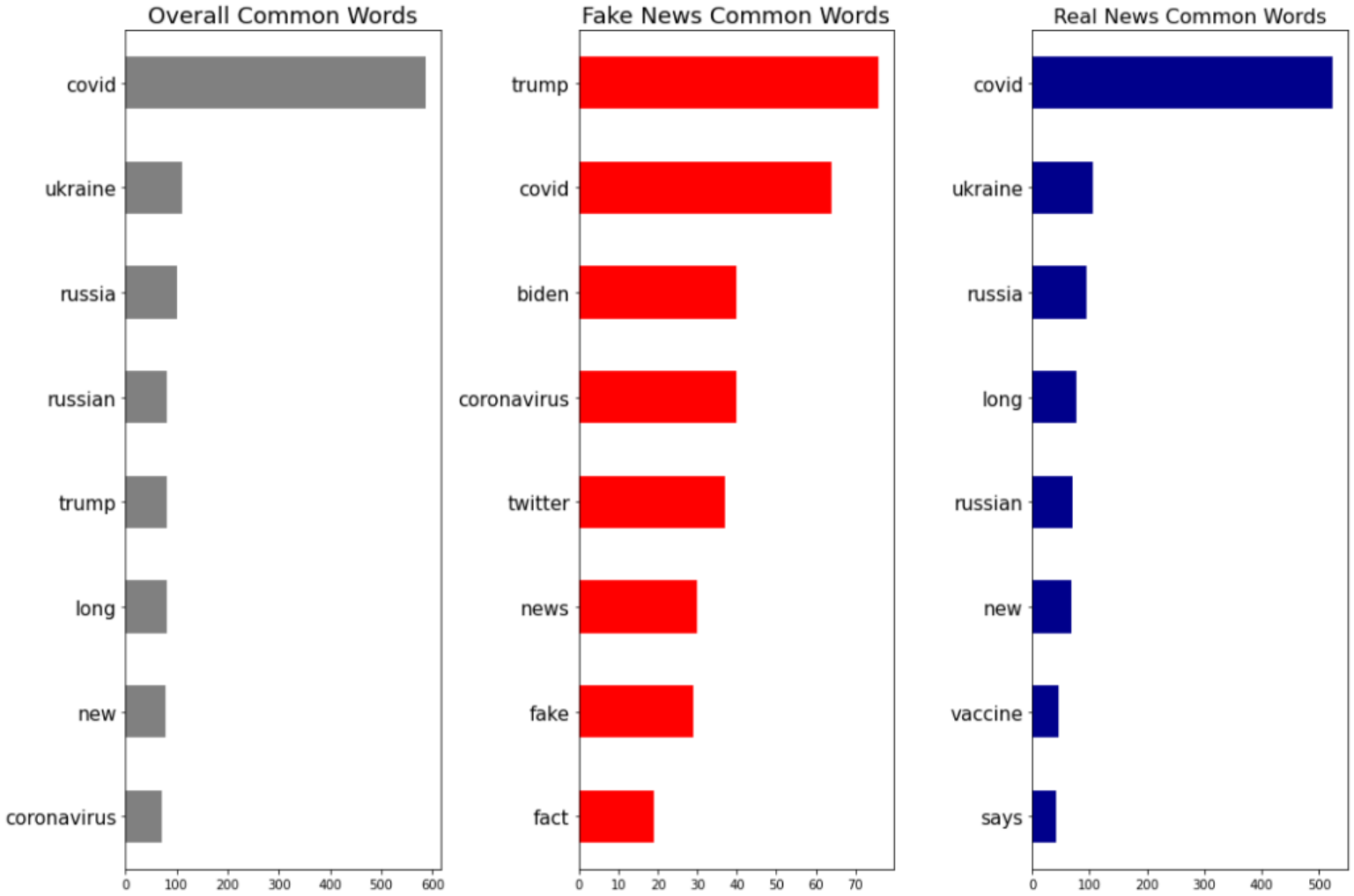
Final Dataset Statistics

The proportion between “Fake News” and “Real News”



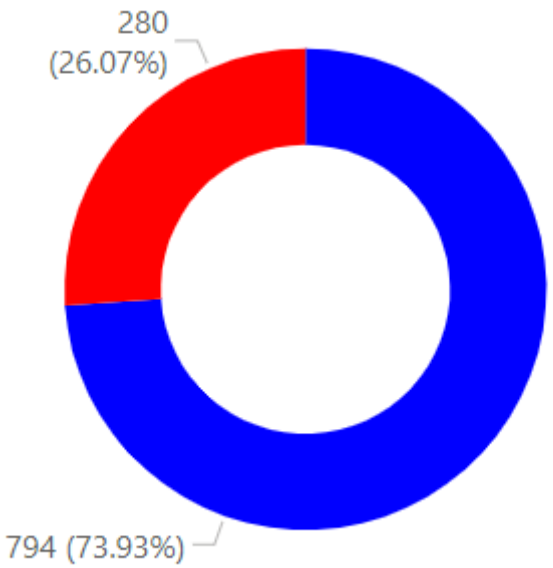
BASELINE SCORE ≈ 74 %

TOP COMMONS WORDS



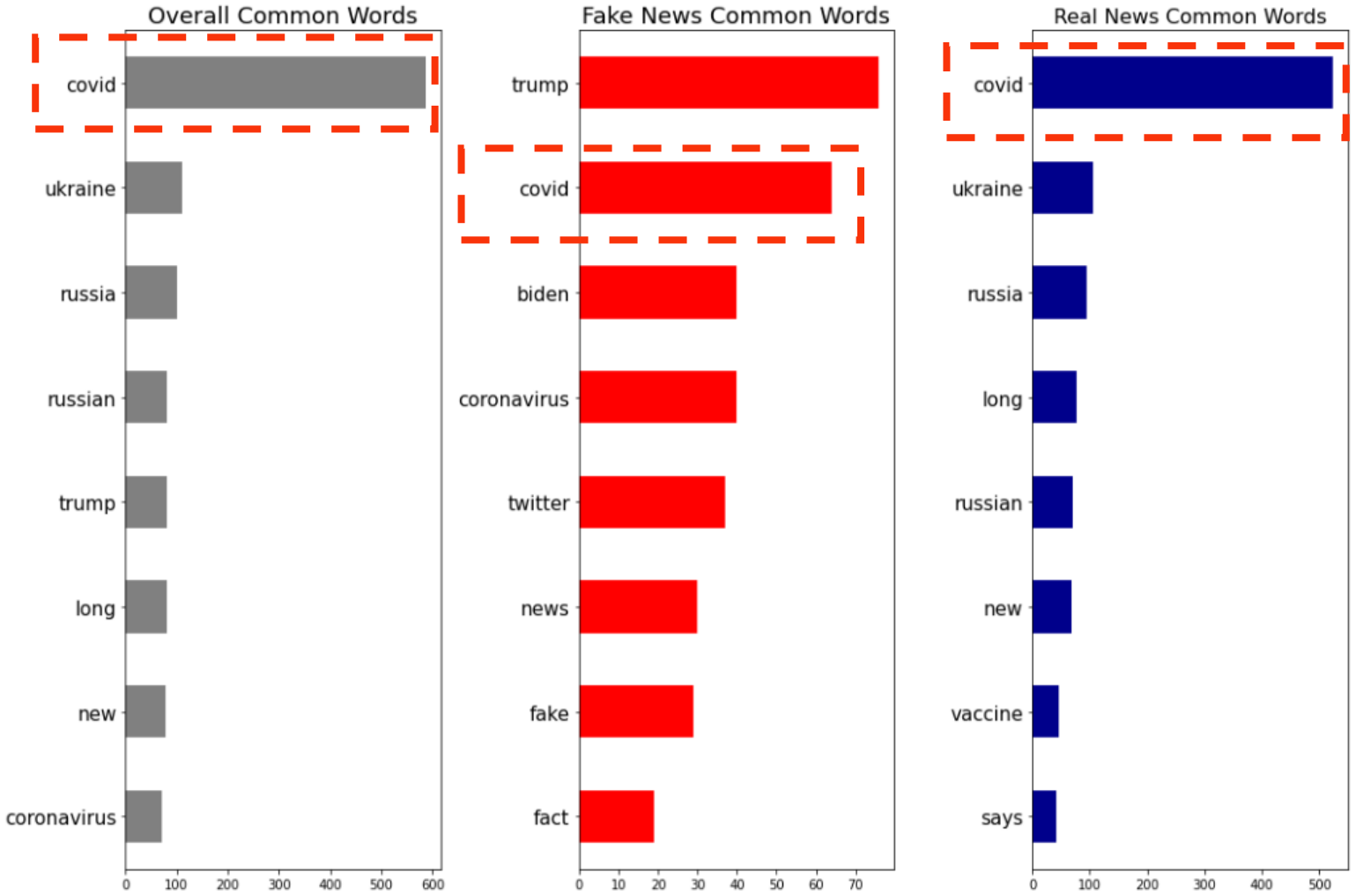
Final Dataset Statistics

The proportion between “Fake News” and “Real News”



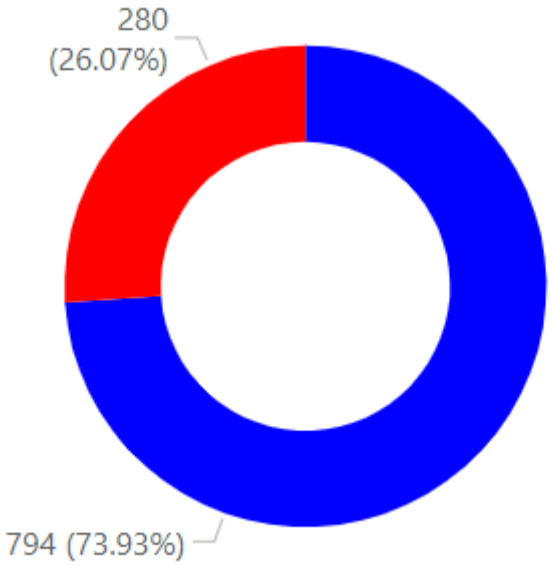
BASELINE SCORE ≈ 74 %

TOP COMMONS WORDS



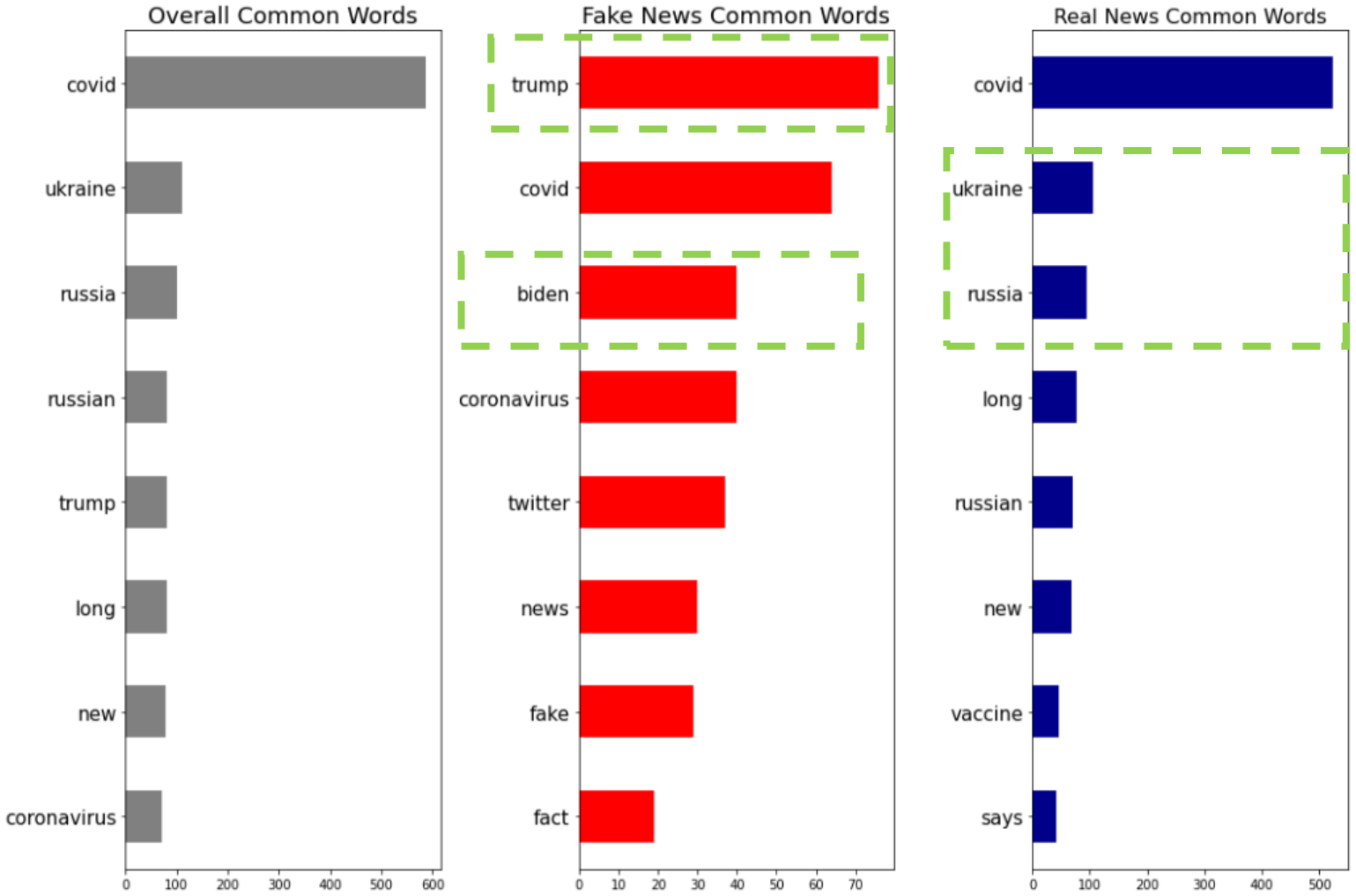
Final Dataset Statistics

The proportion between “Fake News” and “Real News”



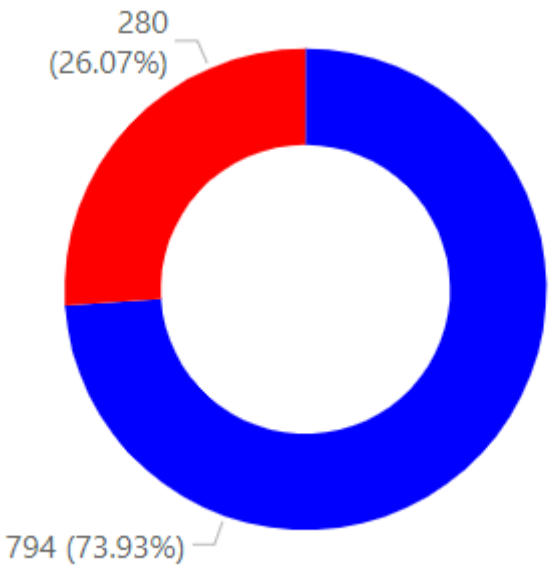
BASELINE SCORE ≈ 74 %

TOP COMMONS WORDS



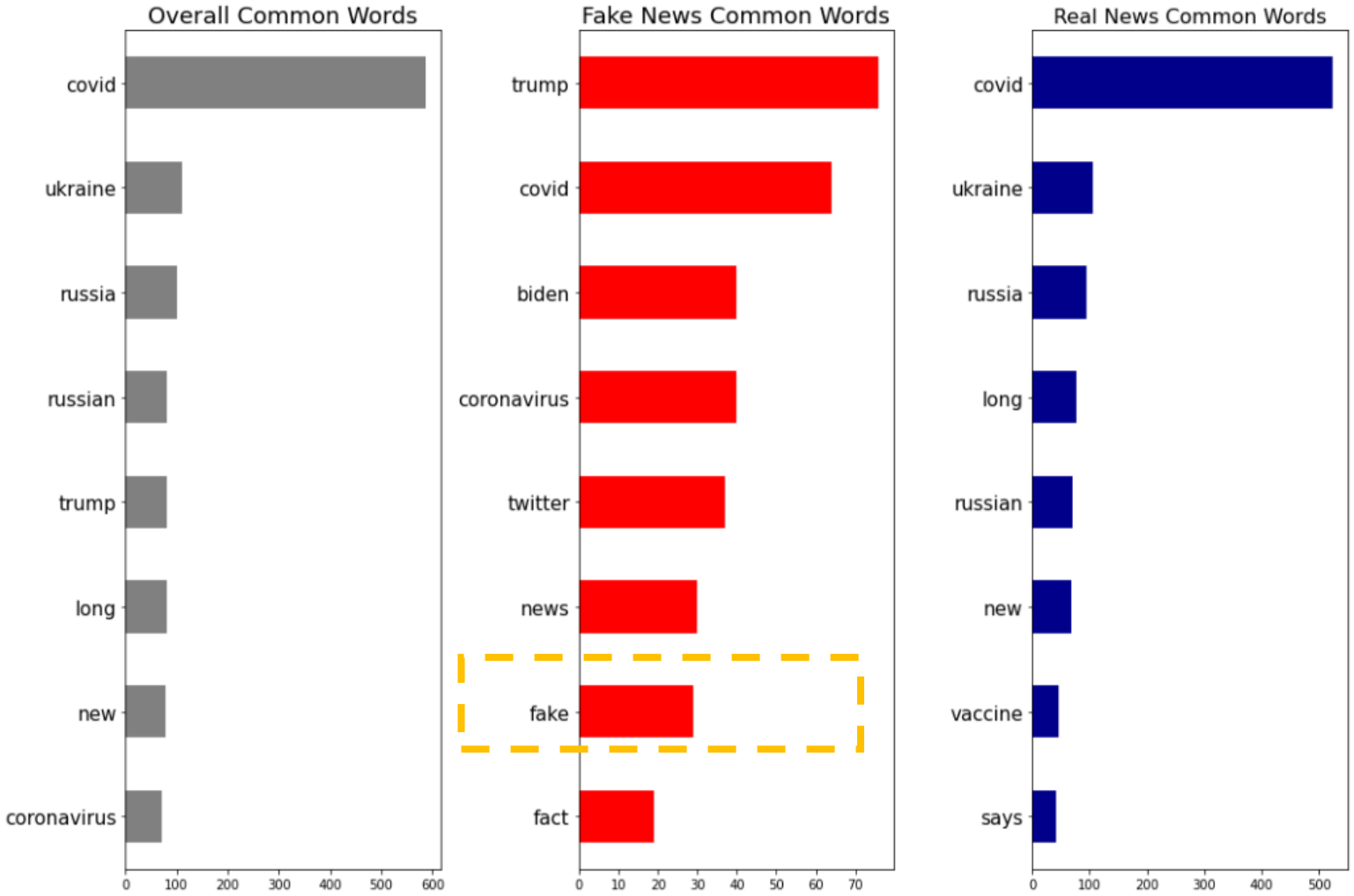
Final Dataset Statistics

The proportion between “Fake News” and “Real News”



BASELINE SCORE ≈ 74 %

TOP COMMONS WORDS

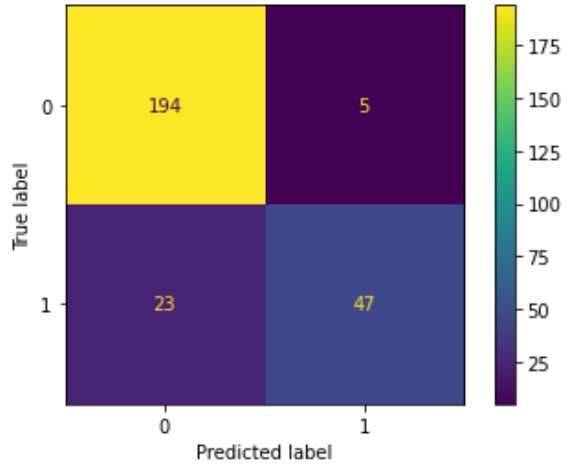


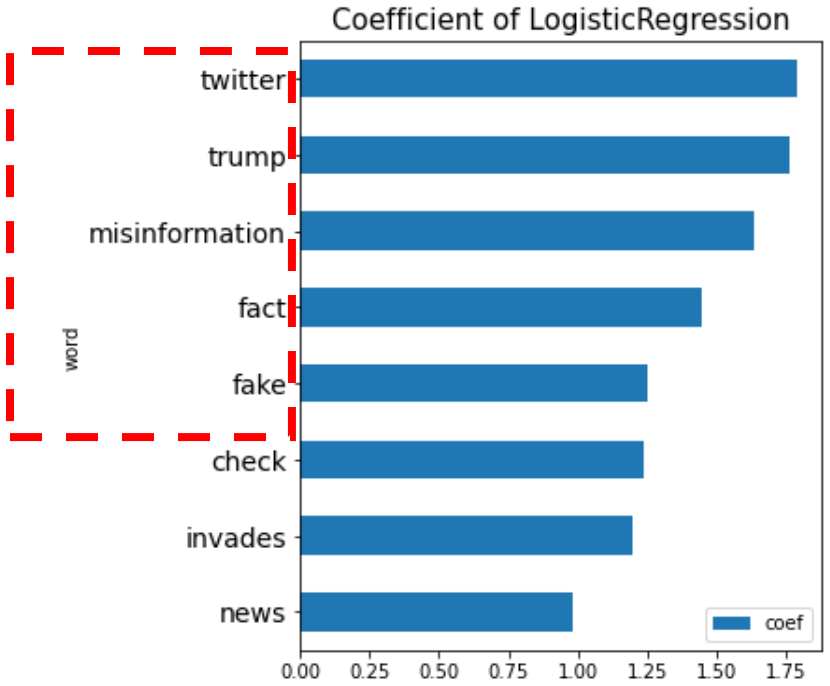
Fake vs. Real News Keys Characteristics

Length	**Fake**		**Real**
Capital Letter	**Fake**		**Real**
Number of Exclamation Marks (!)	**Fake**		**Real**
Noun Per Length	**Fake**		**Real**
Adjective Per Length	**Fake**		**Real**

First Trial-Model

Confusion Matrix





Train Accuracy : 0.96 ***Test Accuracy : 0.90***

False Positive *(Predicted Real as Fake)*

- **Biden** Surprises Elton John With The National Humanities Medal Following A Concert At The White House For His Years Of Advocacy In The Fight Against HIV/AIDS
- **Biden** to pardon all prior federal offenses of simple marijuana possession.
- **Biden** Pardons Thousands of People Convicted of Marijuana Possession Under Federal Law.

False Negative *(Predicted Fake as Real)*

- U.N. Mysteriously Disappears After Criticizing **Russia.**
- Confused **Russian** Soldier Was Told Ukrainians Would Be Happy To Be Summarily Executed In Street
- Biden Addresses **Ukrainian** Crisis With Speech About Perfect Malted Milkshake He Once Drank In 1957

Reliability *Improvement Journey*

1

KEYWORD REMOVAL

Removed Topics word i.e. **Trump, Biden, Putin** etc.

2

PROPER NOUN REMOVAL

Eliminated the specific words i.e. **Elon Musk, Crimea**.

3

CHEAT WORD REMOVAL

Removed the cheated word i.e. **parody, misinformation** etc.

4

IMBALANCED WORD REMOVAL

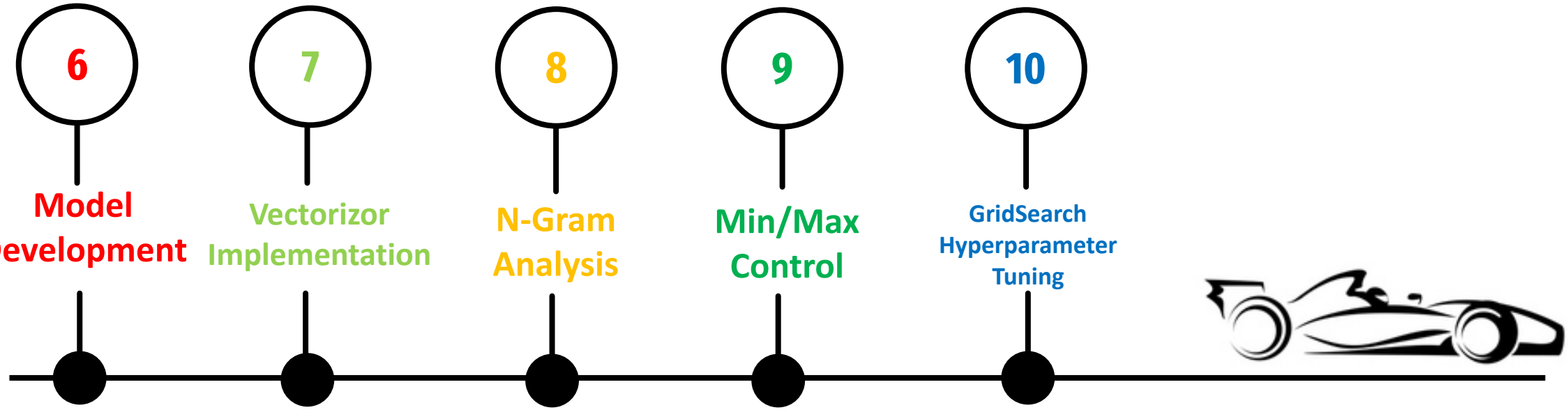
Eliminated the words which has imbalance proportion between class i.e. **Post, Account** etc.

5

STOPWORDS RE-ADDING

Fill the **English** stop words in our tokenized word.

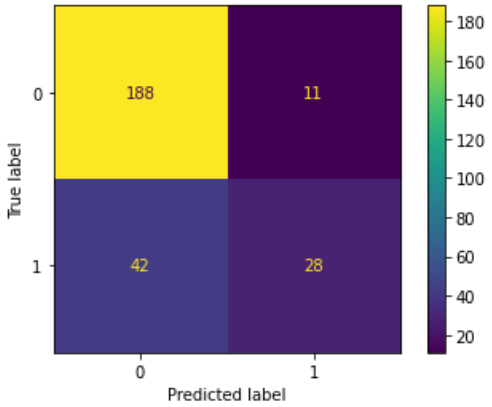
Reliability Improvement Journey (Cont.)



FINAL MODEL SUMMARY

Predicted Model : Multinomial Naïve Bayes

Confusion Matrix



Best Parameter

CountVectorizer

Max Features : 1000

Min DF : 1

N-gram Range : (1,1)

Train Accuracy : 0.92

BASELINE SCORE ≈ 74 %

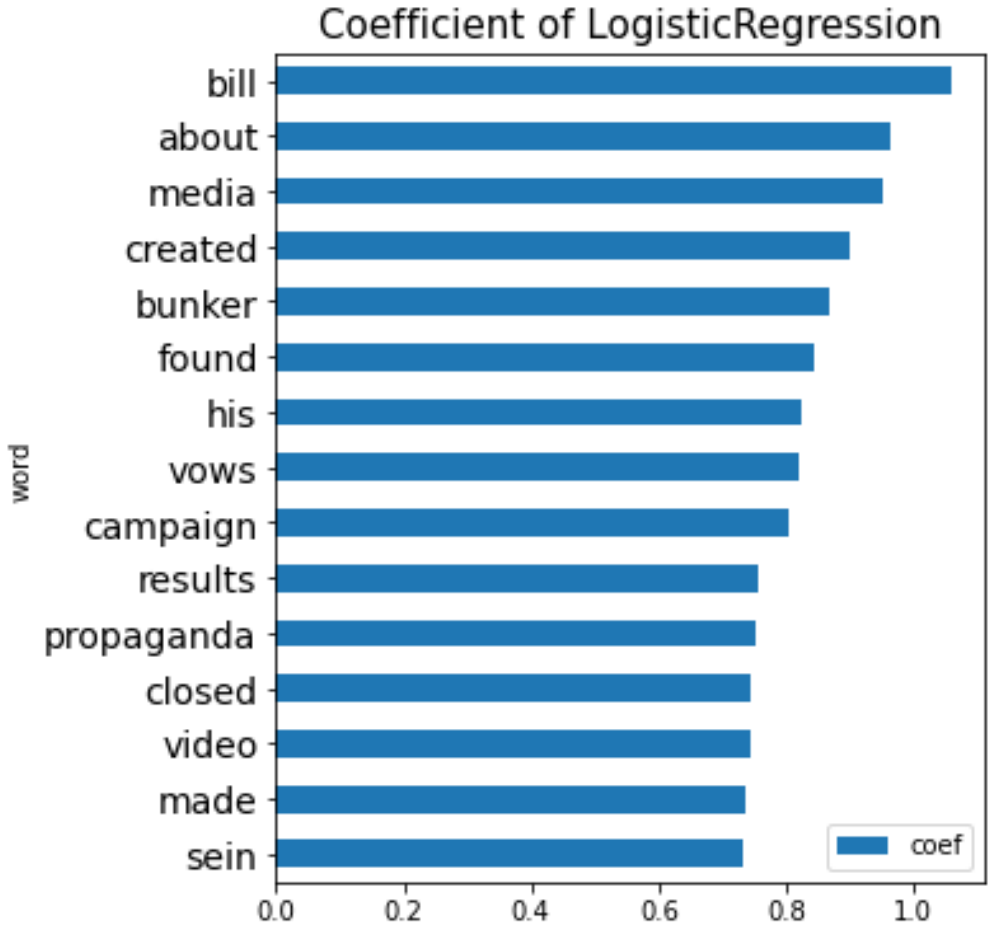
CLASS/METRICS	Precision	Recall	F1-SCORE
REAL	0.92	0.97	0.95
FAKE	0.91	0.77	0.84

Test Accuracy : 0.80

CLASS/METRICS	Precision	Recall	F1-SCORE
REAL	0.82	0.94	0.88
FAKE	0.72	0.4	0.51

FINAL MODEL SUMMARY

Explained Model : Logistic Regression



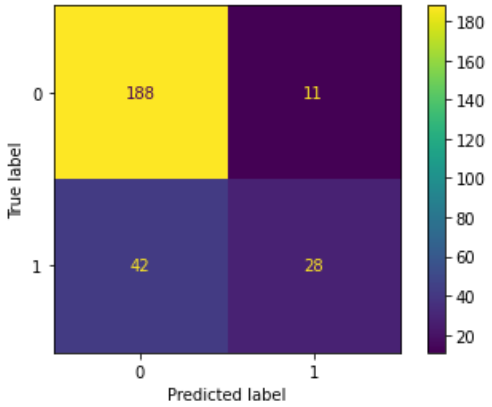
Coefficient to indicate "Fake News Class"

FINAL MODEL SUMMARY

Predicted Model : Multinomial Naïve Bayes

Explained Model : Logistic Regression

Confusion Matrix



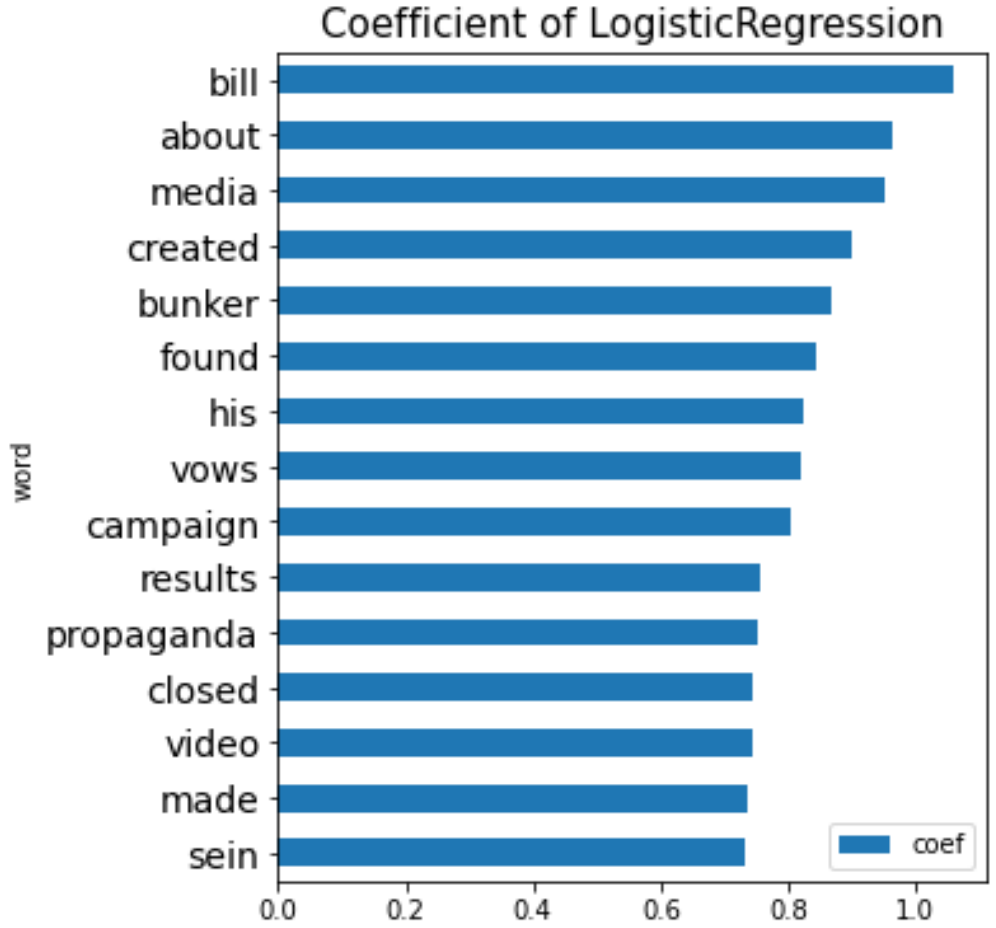
Best Parameter

CountVectorizer

Max Features : 1000

Min DF : 1

N-gram Range : (1,1)



Train Accuracy : 0.92

BASELINE SCORE ≈ 74 %

CLASS/METRICS	Precision	Recall	F1-SCORE
REAL	0.92	0.97	0.95
FAKE	0.91	0.77	0.84

Test Accuracy : 0.80

CLASS/METRICS	Precision	Recall	F1-SCORE
REAL	0.82	0.94	0.88
FAKE	0.72	0.4	0.51

Coefficient to indicate "Fake News Class"

Prediction Results?

Biden sends a careful but chilling new nuclear message to Putin in CNN interview

Analysis by Stephen Collinson, CNN
Updated 6:57 AM EDT, Wed October 12, 2022

MYTH: A video of Ukrainian President Volodymyr Zelensky in a conference call shows cocaine on his desk

Following Russia's full-scale invasion of Ukraine in February 2022, false claims that Ukrainian President Volodymyr Zelensky has a cocaine addiction appeared on pro-Kremlin websites and social media accounts, citing as evidence manipulated or misleading footage.

In late-April 2022, Ilya Kiva, a pro-Russian former Ukrainian Member of Parliament, shared on his Telegram channel a video of a March 2022 conference call between Zelensky and Tesla CEO Elon Musk. Kiva claimed that the video showed white powder and a credit card on Zelensky's desk. Russian website Chelindustry.ru reported on the same video in April 2022, stating: "At

Reference : https://edition.cnn.com/2022/10/12/politics/joe-biden-nuclear-message-putin-cnntv-analysis/index.html

Reference : https://www.newsguardtech.com/special-reports/russian-disinformation-tracking-center.

Prediction Results? - ANSWER

Biden sends a careful but chilling new nuclear message to Putin in CNN interview

Analysis by Stephen Collinson, CNN
Updated 6:57 AM EDT, Wed October 12, 2022

Real News

Correct Prediction

MYTH: A video of Ukrainian President Volodymyr Zelensky in a conference call shows cocaine on his desk

Following Russia's full-scale invasion of Ukraine in February 2022, false claims that Ukrainian President Volodymyr Zelensky has a cocaine addiction appeared on pro-Kremlin websites and social media accounts, citing as evidence manipulated or misleading footage.

In late-April 2022, Ilya Kiva, a pro-Russian former Ukrainian Member of Parliament, shared on his Telegram channel a video of a March 2022 conference call between Zelensky and Tesla CEO Elon Musk. Kiva claimed that the video showed white powder and a credit card on Zelensky's desk. Russian website Chelindustry.ru reported on the same video in April 2022, stating: "At

Fake News

Correct Prediction

Reference : https://edition.cnn.com/2022/10/12/politics/joe-biden-nuclear-message-putin-cnntv-analysis/index.html

Reference : https://www.newsguardtech.com/special-reports/russian-disinformation-tracking-center.

Prediction Results?

Nigeria Feeding Human Corpses to Zoo Animals

Share Tweet G+ +

GOMBE, Nigeria –

As the Boko Haram insurgency increases the death toll of Nigerian Christians, Gombe community leaders are struggling to find money for qualified morticians, caskets and the proper burial of the bodies.

Assange Truth Bomb: Hillary Clinton And ISIS Are Funded By The Same Money

Share Tweet G+ +

Jurassic Marijuana Plant Brought Back To Life After 200 Million Years

Share Tweet G+ +

PARIS, France –

A French team of scientists from the University of Paris fell upon an unlikely find when they gathered remnants of plant material buried deep into the permafrost of Antarctica in 2013. To their surprise, the recovered plant material contained intact seeds, but it is only months later that a team of paleobotanists took up the challenge of reviving the 200 million-year old plant and, to their own disbelief, succeeded.

BirdJR and Gun are President and Vice President of United State

Reference *: http://cityworldnews.com/nigeria-human-corpses-zoo/*

http://cityworldnews.com/jurassic-marijuana-plant/

http://cityworldnews.com/assange-truth-bomb-hillary-clinton-and-isis-are-funded-by-the-same-money/

Prediction Results?

Nigeria Feeding Human Corpses to Zoo Animals

Share Tweet G+

Real News

Incorrect Prediction

GOMBE, Nigeria –

As the Boko Haram insurgency increases the death toll of Nigerian Christians, Gombe community leaders are struggling to find money for qualified morticians, caskets and the proper burial of the bodies.

Jurassic Marijuana Plant Brought Back To Life After 200 Million Years

Share Tweet G+ +

Real News

Incorrect Prediction

PARIS, France –

A French team of scientists from the University of Paris fell upon an unlikely find when they gathered remnants of plant material buried deep into the permafrost of Antarctica in 2013. To their surprise, the recovered plant material contained intact seeds, but it is only months later that a team of paleobotanists took up the challenge of reviving the 200 million-year old plant and, to their own disbelief, succeeded.

Assange Truth Bomb: Hillary Clinton And ISIS Are Funded By The Same Money

Share Tweet G+ +

Real News

Incorrect Prediction

BirdJR and Gun are President and Vice President of United State

Real News

Incorrect Prediction

Reference : http://cityworldnews.com/nigeria-human-corpses-zoo/
http://cityworldnews.com/jurassic-marijuana-plant/
http://cityworldnews.com/assange-truth-bomb-hillary-clinton-and-isis-are-funded-by-the-same-money/

STRENGTH

- Have applicable level of Accuracy, Precision and Recall for Fake News Detection in Reddit News with learned topics.

WEAKNESS

- Less predictability in unlearned topics, data source and platforms.
- High bias in the difference topic proportion between Real and Fake News.
- Unable to predict an AI-generated or High similarity Fake News.
- Unable to detect the key properties of Documents i.e. Structure, Position etc.

WAY FORWARD

more subreddit

Diversify Data Source

To explore multiple Fake and Real News data sources not only on reddit but on other platforms.

Increase Data Size

To gather more datasets to potentially increase model performance and reduce overfitting.

Explore Higher Model Predictability

To explore more predictively powerful algorithm, especially in **BERT-based** algorithms to maximize model performance.

Consult Domain Experts

To seek for the Fake News Subject Matter Expert **(SME)** to debottleneck of Fake News Elimination Projects.


QUESTIONS
Q&A
ANSWERS

Who are we ?

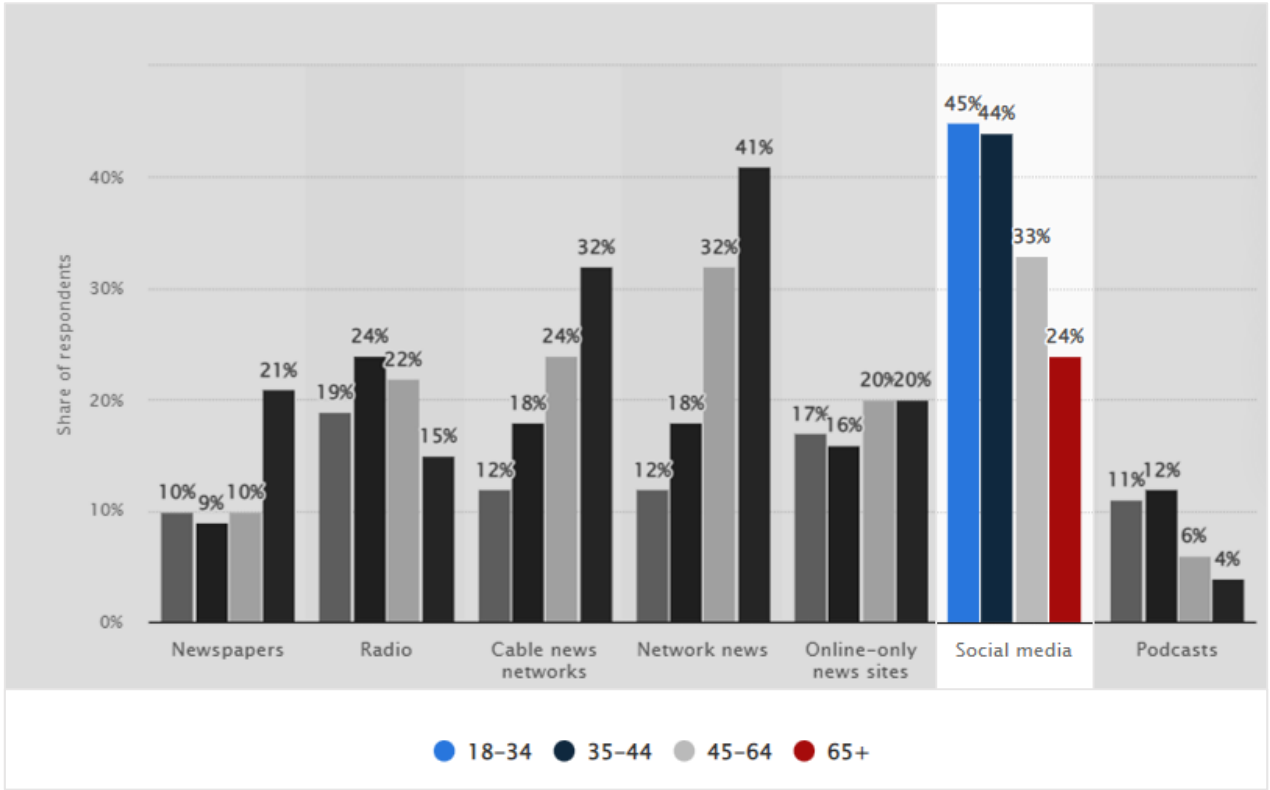
We are ***Central Data Scientist Team*** of **ADVANCE PUBLICATIONS GROUP** who have been assigned to algorithm to remove “Fake News” in Reddit accord ***Reddit IPO Readiness Plan*** in 2022.

REDDIT IPO
READINESS PLAN 2022

Our Journey

Reference : https://www.statista.com/topics/3251/fake-news/#topicHeader__wrapper

Most popular platforms for daily news consumption in the United States as of February 2022, by age group



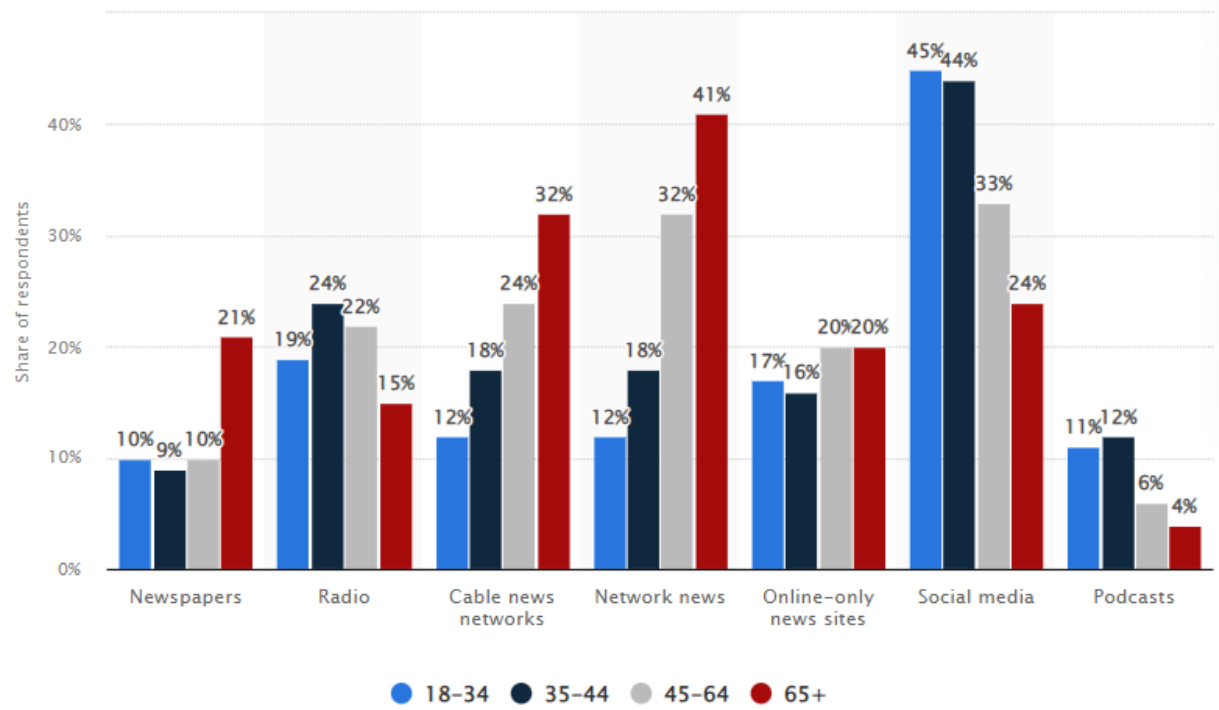
PROJECT 3

FAKE NEWS DETECTION IN REDDIT NEWS DATABASE

Chalermchon Wongsopa

Warintorn Nawong

Most popular platforms for daily news consumption in the United States as of February 2022, by age group



Additional Information

Show source

DOWNLOAD

PDF XLS PNG PPT

Source

- Show sources information
- Show publisher information
- Use Ask Statista Research Service

Release date

February 2022

Region

United States

Survey time period

February 9 to 10, 2022

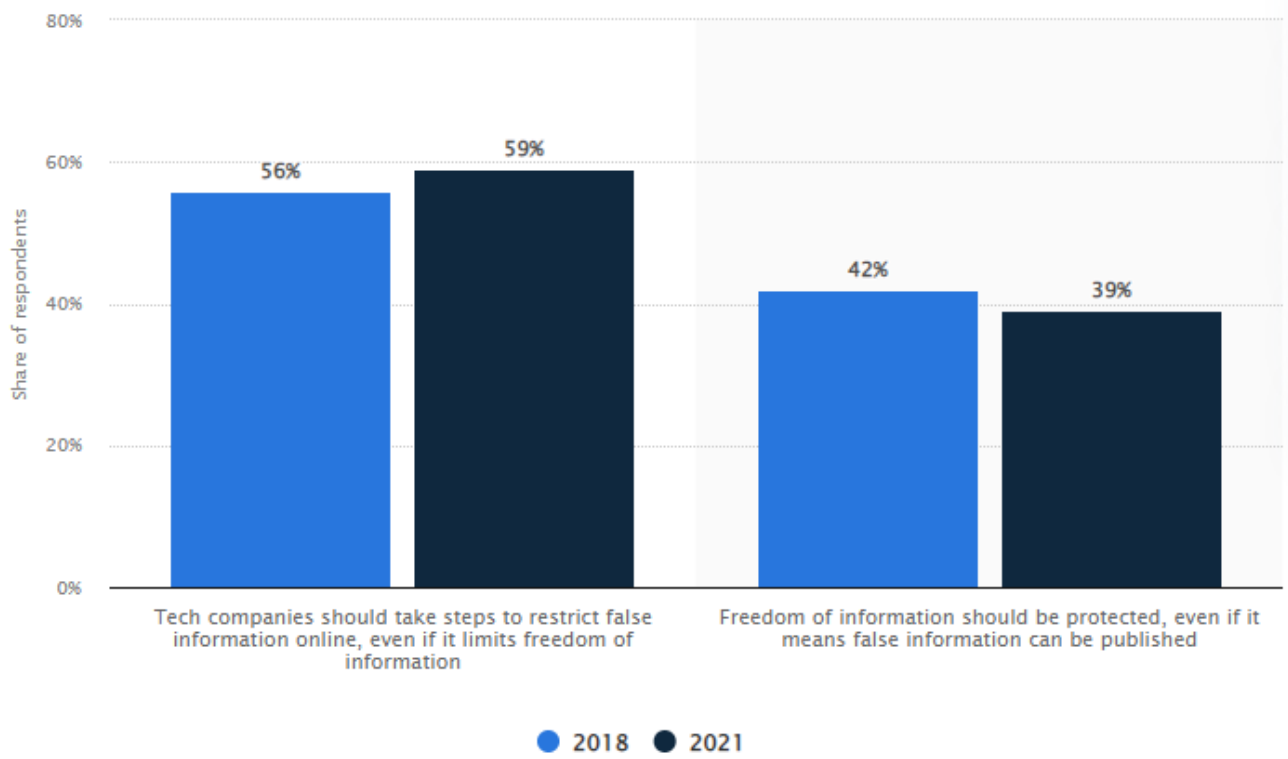
Number of respondents

2,210 respondents

Age group

18 years and older

Public opinion on tech companies restricting fake news online in the United States in 2018 and 2021



Additional Information

Show source

DOWNLOAD

PDF + XLS + PNG + PPT +

Source

→ Show sources information

→ Show publisher information

→ Use Ask Statista Research Service

Release date

August 2021

Region

United States

Survey time period

February 26 to March 11, 2018 and July 26 to August 8, 2021

Number of respondents

2018: 4,734; 2020: 11,178

Age group

18 years and older

Fake news in the U.S. - statistics & facts

Published by Amy Watson, Jun 21, 2022

Fake news is an insidious and widespread issue in the news industry as a whole and has become a global problem. In the United States, the term and concept grew in popularity during the 2016 election, but has since manifested itself in areas outside the realm of politics. A recent example of this is the COVID-19 pandemic – almost 80 percent of consumers in the United States reported having seen fake news on the coronavirus outbreak, highlighting the extent of the issue and the reach fake news can achieve.

Read more

STATISTICS ON THE TOPIC

SHARE OF AMERICANS VERY CONFIDENT IN THEIR ABILITY TO RECOGNIZE FAKE NEWS

26%

AMERICANS WHO BELIEVE FAKE NEWS CAUSES A GREAT DEAL OF CONFUSION

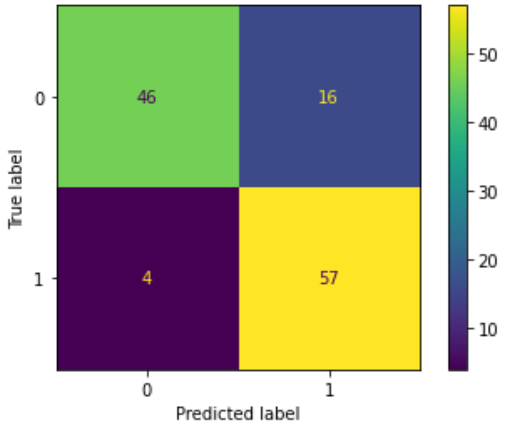
67%

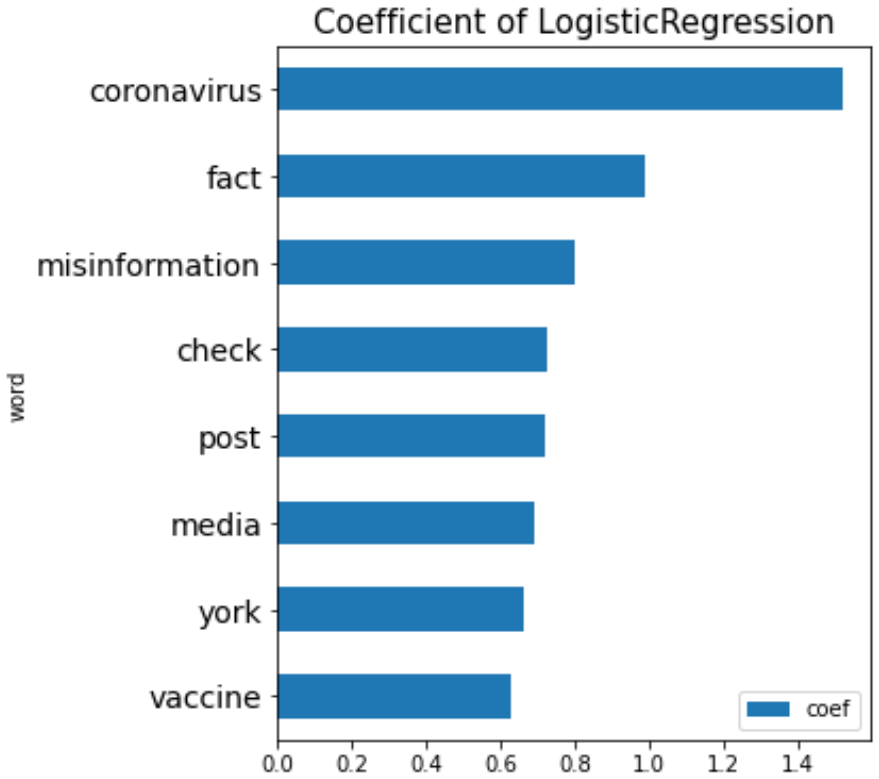
AMERICANS WHO ACCIDENTALLY SHARED FAKE NEWS

38.2%

First Trial-Model.

Confusion Matrix





Train Score : 1.00 ***Test Score : 0.84***

False Positive.

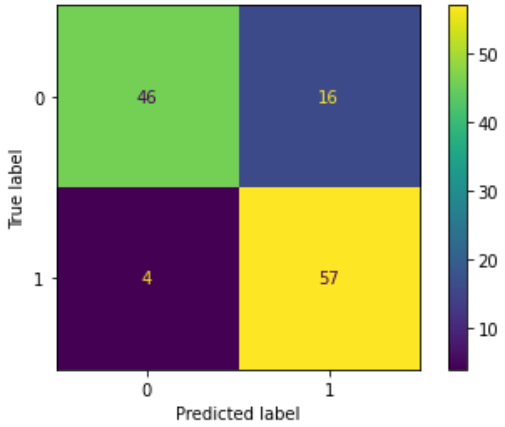
- **Trump** asks U.S. Supreme Court to intervene over seized classified records.
- **Biden** Addresses Ukrainian Crisis With Speech About Perfect Malted Milkshake He Once Drank In 1957.
- China holiday tourist trips fall 18% on year on broad **COVID** curbs.

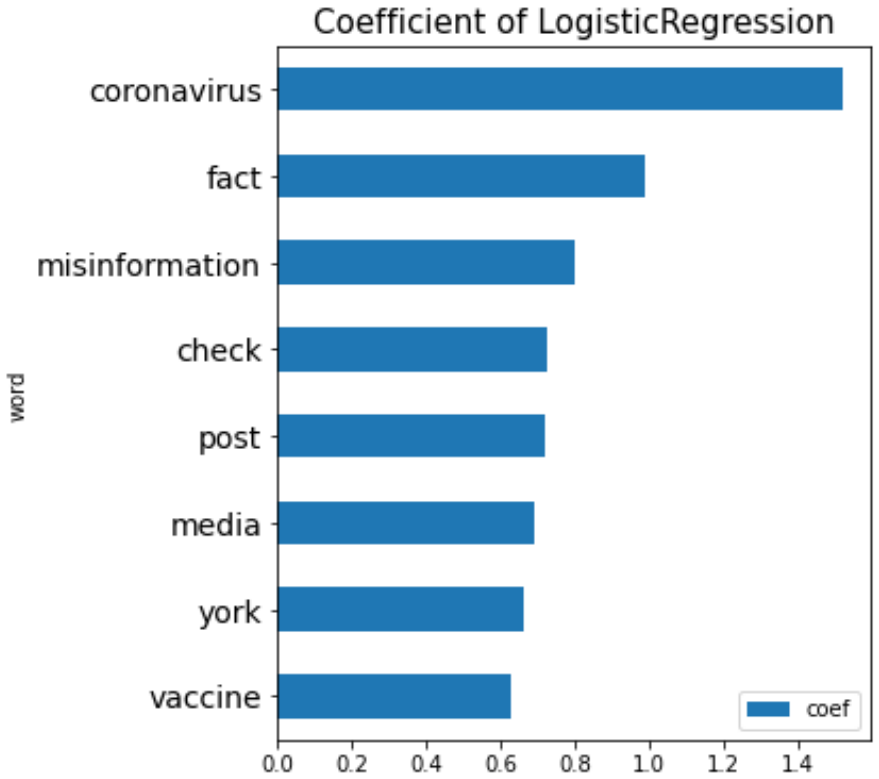
False Negative.

- **Russian** Soldiers Guns, Tanks Vanish Into Thin Air As First Wave Of Sanctions Takes Effect.
- From **Russia** with Love: social networks and Russian outlets as a tool to promote Sputnik V in Latin America
- A Bible Burning, a **Russian** News Agency and a Story Too Good to Check Out

First Trial-Model.

Confusion Matrix





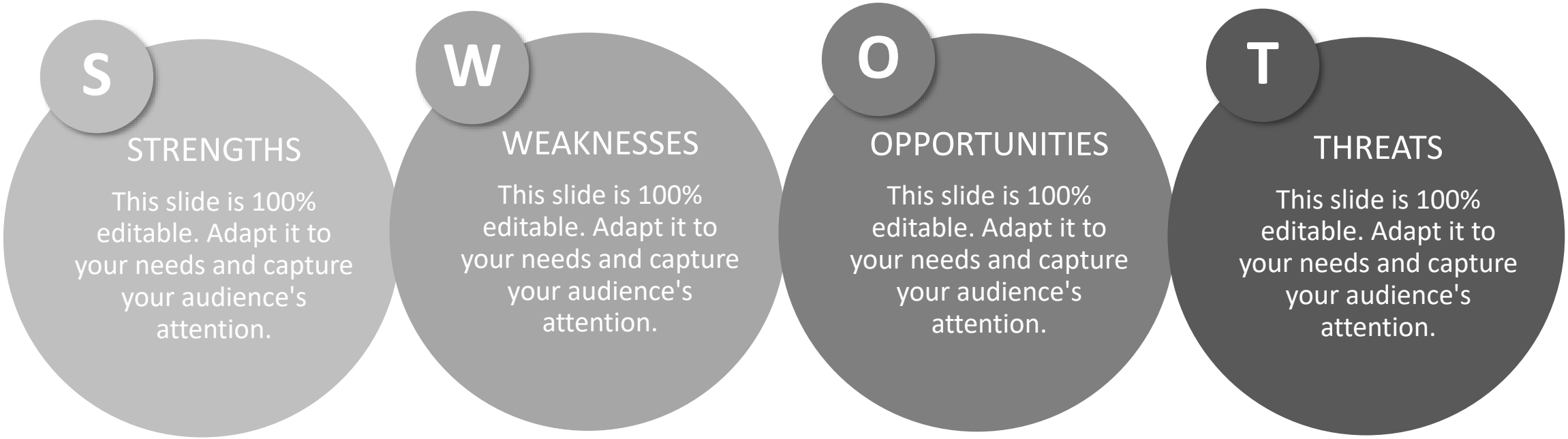
Train Score : 1.00 ***Test Score : 0.84***

False Positive.

- **Trump** asks U.S. Supreme Court to intervene over seized classified records.
- **Biden** Addresses Ukrainian Crisis With Speech About Perfect Malted Milkshake He Once Drank In 1957.
- China holiday tourist trips fall 18% on year on broad **COVID** curbs.

False Negative.

- **Russian** Soldiers Guns, Tanks Vanish Into Thin Air As First Wave Of Sanctions Takes Effect.
- From **Russia** with Love: social networks and Russian outlets as a tool to promote Sputnik V in Latin America
- A Bible Burning, a **Russian** News Agency and a Story Too Good to Check Out


S
STRENGTHS
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.
W
WEAKNESSES
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.
O
OPPORTUNITIES
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.
T
THREATS
This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Vehicles Infographics

1

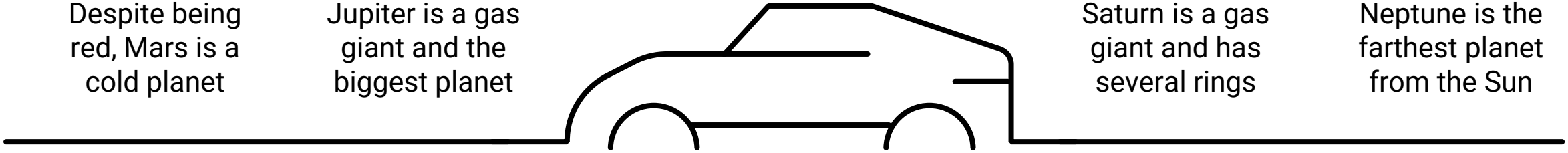
Mars

Despite being red, Mars is a cold planet

2

Jupiter

Jupiter is a gas giant and the biggest planet



3

Saturn

Saturn is a gas giant and has several rings

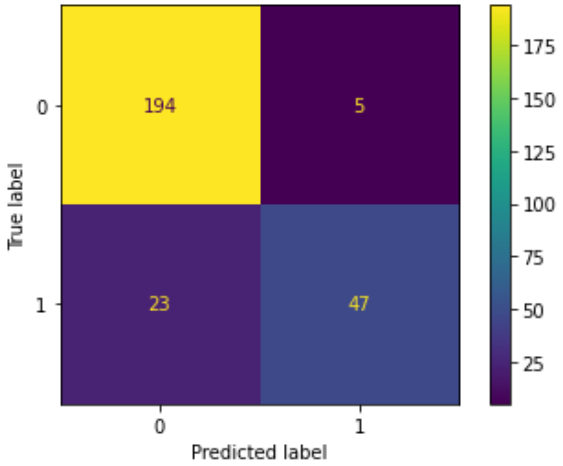
4

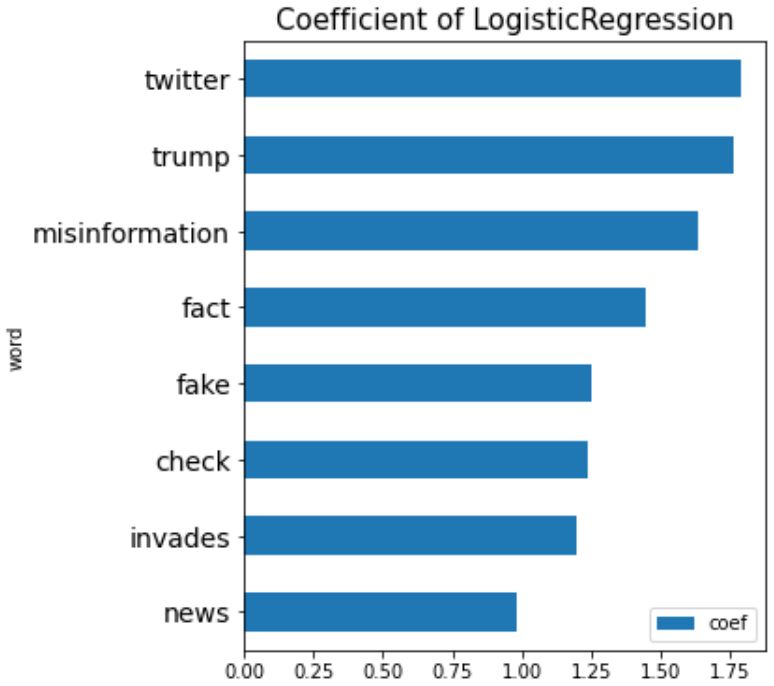
Neptune

Neptune is the farthest planet from the Sun

First Trial-Model.

Confusion Matrix





Train Accuracy : **0.96** *Test Accuracy :* **0.90**