



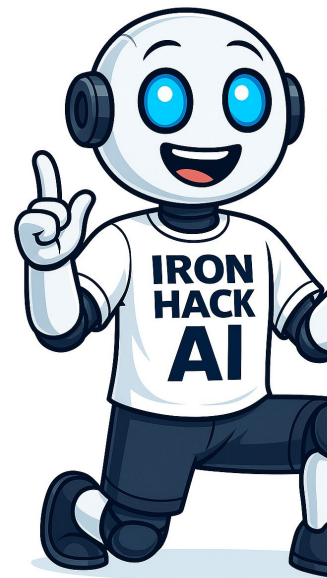
IRON - HACK NLP PROJECT - August 2025

LUKMAN SADIQ

PROBLEM STATEMENT - Deep Learning: Image Classification with CNN

To - Natural Language Processing Challenge

DATASET



DATA

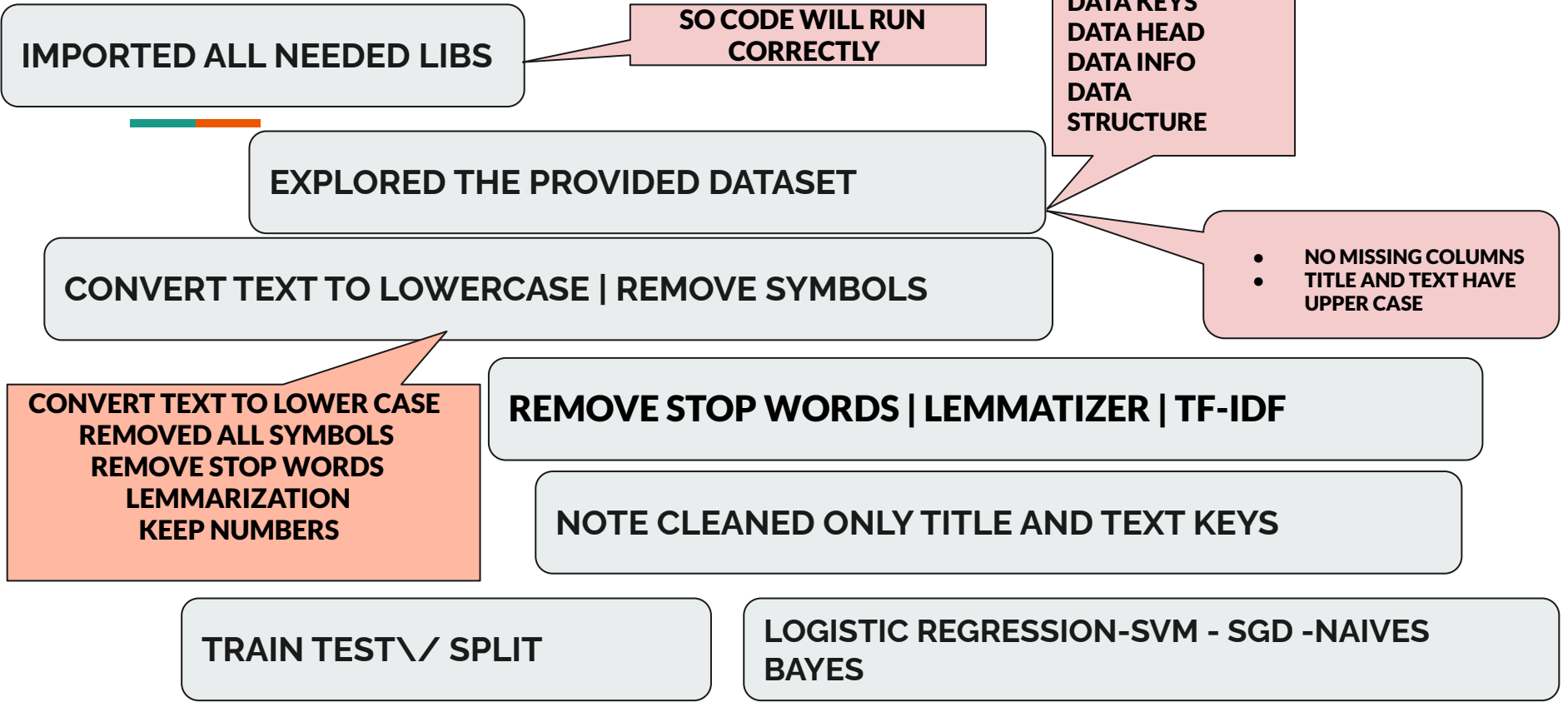
CIFAR - 10 DATA INFO- DISTRIBUTION LEARNINGS

```
label      0
title      0
text       0
subject    0
date       0
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39942 entries, 0 to 39941
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   label       39942 non-null  int64
1   title       39942 non-null  object
2   text        39942 non-null  object
3   subject     39942 non-null  object
4   date        39942 non-null  object
dtypes: int64(1), object(4)
memory usage: 1.5+ MB
```

label		title	text	subject	date
0	2	UK's May 'receiving regular updates' on London...	LONDON (Reuters) - British Prime Minister Ther...	worldnews	September 15, 2017
1	2	UK transport police leading investigation of L...	LONDON (Reuters) - British counter-terrorism p...	worldnews	September 15, 2017
2	2	Pacific nations crack down on North Korean shi...	WELLINGTON (Reuters) - South Pacific island na...	worldnews	September 15, 2017
3	2	Three suspected al Qaeda militants killed in Y...	ADEN, Yemen (Reuters) - Three suspected al Qae...	worldnews	September 15, 2017
4	2	Chinese academics prod Beijing to consider Nor...	BEIJING (Reuters) - Chinese academics are publ...	worldnews	September 15, 2017

NLP - WORKFLOW WALKTHROUGH



TRAINED - DATASET *USING* **LOGISTIC REGRESSION**

```
▼ LogisticRegression  
LogisticRegression(max_iter=2000)
```



LogisticRegression

```
LogisticRegression(max_iter=2000)
```

... Accuracy: 0.9875

Classification Report:

	precision	recall	f1-score	support
0	0.989	0.986	0.987	3989
1	0.986	0.989	0.988	4000
accuracy			0.987	7989
macro avg	0.987	0.987	0.987	7989
weighted avg	0.987	0.987	0.987	7989

Confusion Matrix:

```
[[3932  57]
 [  43 3957]]
```

```
# Natural Language Processing Challenge (Complement Naive Bayes)
```

Accuracy: 0.9415

Classification Report:

	precision	recall	f1-score	support
0	0.943	0.940	0.941	3989
1	0.940	0.943	0.942	4000
accuracy			0.942	7989
macro avg	0.942	0.942	0.942	7989
weighted avg	0.942	0.942	0.942	7989

Confusion Matrix:

```
[[3749  240]
 [  227 3773]]
```

```
['tfidf_vectorizer.joblib']
```

TRAINED - DATASET *USING* **SVM**



```
# Using (SVM)
```

Accuracy: 0.995

Classification Report:

	precision	recall	f1-score	support
0	0.995	0.995	0.995	3989
1	0.995	0.995	0.995	4000
accuracy			0.995	7989
macro avg	0.995	0.995	0.995	7989
weighted avg	0.995	0.995	0.995	7989

Confusion Matrix:

```
[[3970  19]
 [  21 3979]]
```

```
# (SGDClassifier)
```

Accuracy: 0.9817

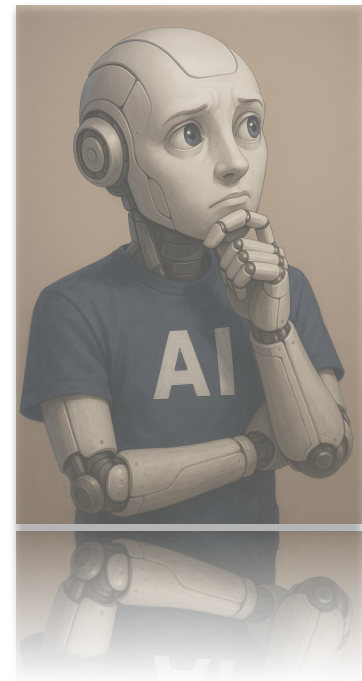
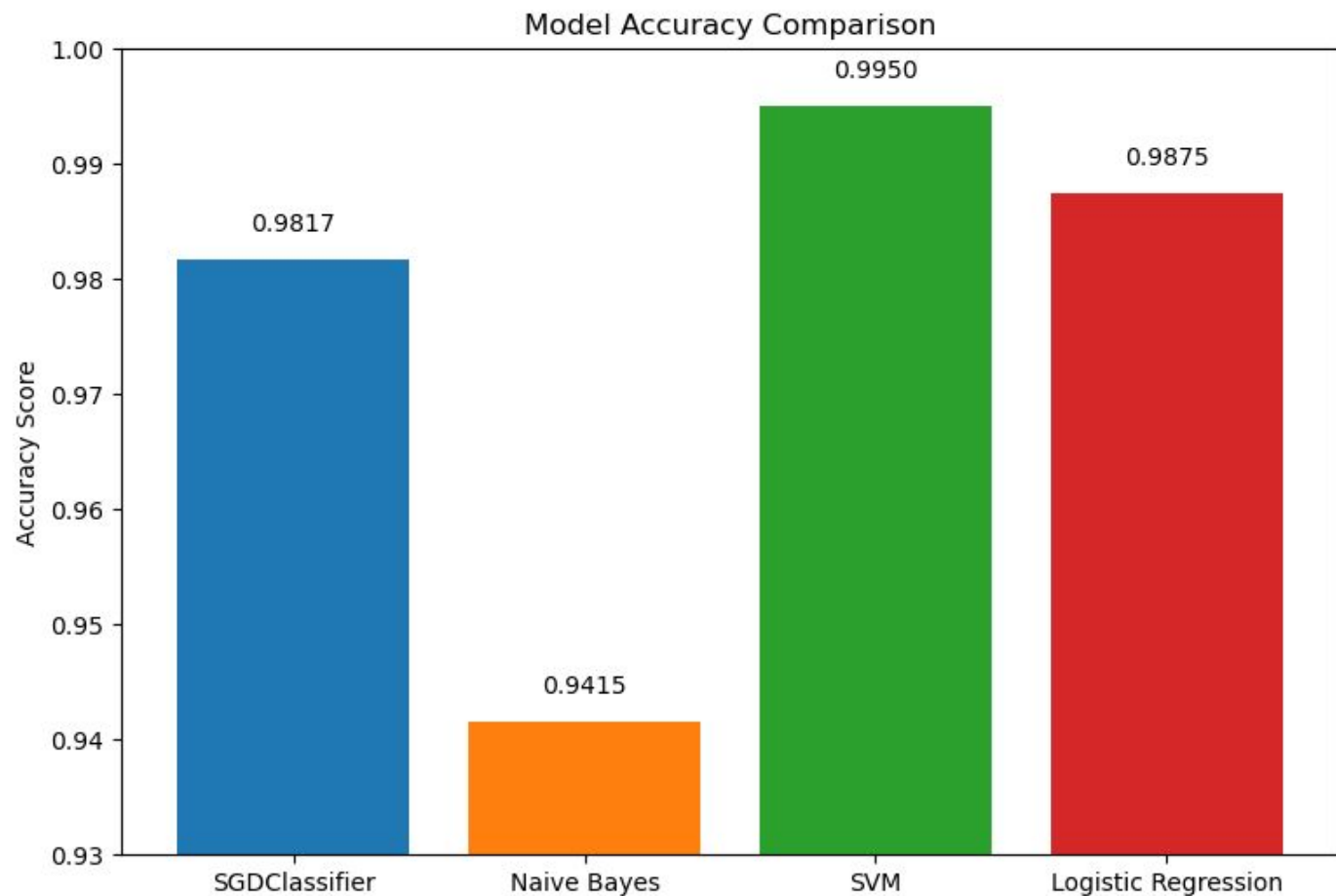
Classification Report:

	precision	recall	f1-score	support
0	0.985	0.978	0.982	3989
1	0.979	0.985	0.982	4000
accuracy			0.982	7989
macro avg	0.982	0.982	0.982	7989
weighted avg	0.982	0.982	0.982	7989

Confusion Matrix:

```
[[3903  86]
 [  60 3940]]
```


PERFORMANCE SUMMARY - SVM BEST SCORE



TESTING TRAINED - MODEL *using* - LINEAR SVC

Marubeni
Institute
株式会社丸紅経済研究所

日本語 English A A

[Company Information](#) [List of reports](#) [Dispatches from the Potomac](#)

Dispatches from the Potomac

#19 | Real Events Caused by Fake News in the US

CNN World [Africa](#) [Americas](#) [Asia](#) [Australia](#) [More](#) ▾

[Watch](#) [Listen](#) [Live TV](#) [Q](#)

[Sign In](#)

WORLD > INDIA • 4 MIN READ

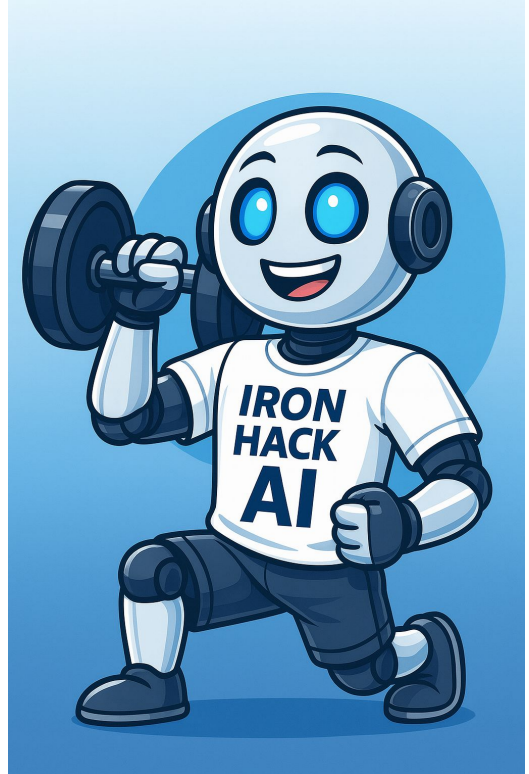
Sudden, heavy rain in Pakistan, India-administered Kashmir and Nepal kills more than 400 people

⚠ Prediction: Fake

Model: LinearSVC • Features: TF-IDF (unigrams+bigrams)

✅ Prediction: Real

Model: LinearSVC • Features: TF-IDF (unigrams+bigrams)





Fake-vs-Real News (SVM)

Paste a headline and/or article text. Uses your saved LinearSVC + TF-IDF vectorizer.

Headline (optional)

Article text (paste here)

campaign. The operators of the pizza shop began receiving threats from right wing activists who believed that the reports were true.

According to The Washington Post, the trigger was the announcement on October 28 about the resumption of the investigation by the Federal Bureau of investigation ("FBI") into the issue of the use of private e-mail by Hillary Clinton during her tenure as Secretary of State. Two days later, there was a large volume of tweets about newly-discovered e-mail related to a pedophile sex ring. These posts were spread on anonymous bulletin board sites and via social media; and there were many abusive videos directed at Ms. Clinton posted on extreme-right-wing sites. Two days before the election, the FBI announced that they had reached a conclusion that they would not seek prosecution of Ms. Clinton.

Classify



Prediction: Fake

Model: LinearSVC • Features: TF-IDF (unigrams+bigrams)

Why? (top weighted words from this text)

Top words pushing toward Real:

shop (0.192), sex (0.091), clinton (0.084), federal bureau (0.082), posted (0.081), washington (0.076), bureau investigation (0.066), democratic (0.061), social medium (0.052), investigation (0.051)

Top words pushing toward Fake:

hillary (-0.266), via (-0.230), video (-0.163), two day (-0.100), announced (-0.087), washington post (-0.077), threat (-0.077), post (-0.060), fbi (-0.060), announcement (-0.054)

Model details

```
{
  "classifier": "LinearSVC"
  "decision_score": -0.7762679750318017
  "vectorizer": "TfidfVectorizer"
  "ngram_range": [
    0 : 1
    1 : 2
  ]
  "max_features": 5000
}
```

Choose a model

SVC



Headline (optional)

Sudden, heavy rain in Pakistan, India-administered Kashmir and Nepal kills more than 400 people

Article text (paste here)

In India-administered Kashmir, at least 60 people died and more than 200 were missing in the town of Chashoti on Friday, a popular pilgrimage destination for Hindu tourists, Reuters reported.

At least 41 people died in Nepal, with another 121 left injured, Reuters reported, citing the country's disaster management authority.

"When the rain intensified, it wasn't long before I felt as if an earthquake had struck — the whole ground was shaking," said Farhad Ali, a student living in Salarzai, in hard-hit northern Pakistan.



Classify



Prediction: Real



Prediction: Real

Model: SVC • Features: TF-IDF (unigrams+bigrams)

Why? (top weighted words from this text)



Top words pushing toward Real:

reuters (0.944), said (0.237), pakistan (0.229), killed (0.214), died (0.135), 400 (0.088), local (0.076), friday (0.068), citing (0.066), authority (0.064)

Top words pushing toward Fake:

prevented (-0.083), living (-0.081), cnn (-0.046), nothing (-0.043), according (-0.042), spokesperson (-0.040), left (-0.036), sudden (-0.035), disaster (-0.030), felt (-0.030)

Model details



Useful - Learnings

 Research on the project is very important.

Using text and title gave good results.

Context is king.