**FAKE NEWS DETECTION USING NLP**

Naive Bayes Classifier Algorithm:

This probabilistic algorithm is often applied to text classification tasks and can help distinguish between real and fake news articles based on the likelihood of certain words or phrases

important libraries:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import nltk

import re

import string

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import classification\_report

import keras

from keras.preprocessing import text,sequence

from keras.models import Sequential

from keras.layers import Dense,Embedding,LSTM,Dropout

import warnings

warnings.filterwarnings('ignore')

import os

for dirname, \_, filenames in os.walk('/kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))

Load and check data:

**in[1]:**

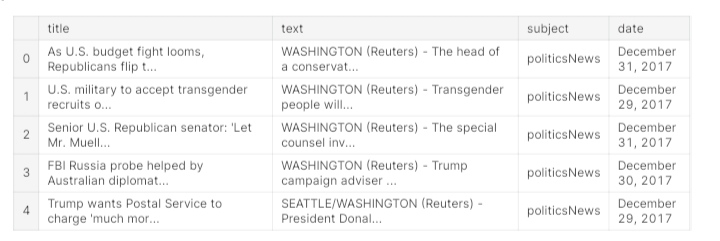
real\_data = pd.read\_csv('/kaggle/input/fake-and-real-news-dataset/True.csv')

fake\_data = pd.read\_csv('/kaggle/input/fake-and-real-news-dataset/Fake.csv')

**in[2]:**

real\_data.head()

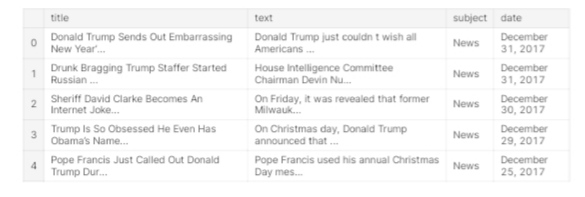
**out[2]:**



**in[3]:**

fake\_data.head()

**out[3]:**



**in[4]:**

**#add column**

real\_data['target'] = 1

fake\_data['target'] = 0

**in[5]:**

real\_data.tail()

**out[5]:**



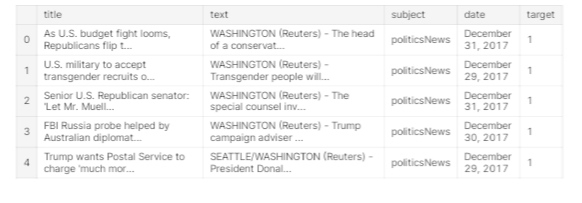
**in[6]:**

**#Merging the 2 datasets**

data = pd.concat([real\_data, fake\_data], ignore\_index=True, sort=False)

data.head()

**out[6]:**



**in[7]:**

data.isnull().sum()

**out[7]:**

title 0

text 0

subject 0

date 0

target 0

dtype: int64

Visualization:

1.Count of Fake and Real Data:

**in[8]:**

fig, ax = plt.subplots(1,2, figsize=(19, 5))

g1 = sns.countplot(data.target,ax=ax[0],palette="pastel");

g1.set\_title("Count of real and fake data")

g1.set\_ylabel("Count")

g1.set\_xlabel("Target")

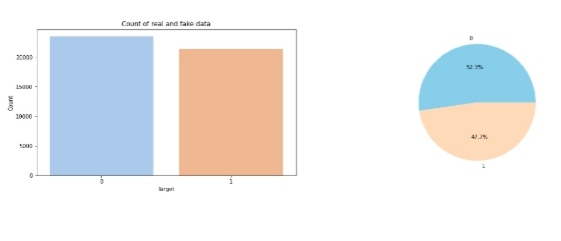
g2 = plt.pie(data["target"].value\_counts().values,explode=[0,0],labels=data.target.value\_counts().index, autopct='%1.1f%%',colors=['SkyBlue','PeachPuff'])

fig.show()

**0 23481**

**1 21417**

**Name: target, dtype: int64**



2.Distribution of The Subject According to Real and Fake Data:

**in[9]:**

print(data.subject.value\_counts())

plt.figure(figsize=(10, 5))

ax = sns.countplot(x="subject", hue='target', data=data, palette="pastel")

plt.title("Distribution of The Subject According to Real and Fake Data")

**politicsNews 11272**

**worldnews 10145**

**News 9050**

**politics 6841**

**left-news 4459**

**Government News 1570**

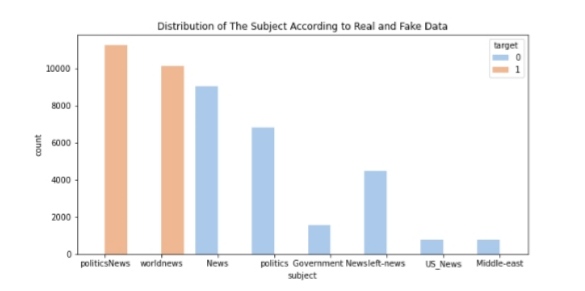
**US\_News 783**

**Middle-east 778**

**Name: subject, dtype: int64**

**out[9]:**

Text(0.5, 1.0, 'Distribution of The Subject According to Real and Fake Data')



Data Cleaning:

**in[10]:**

data['text']= data['subject'] + " " + data['title'] + " " + data['text']

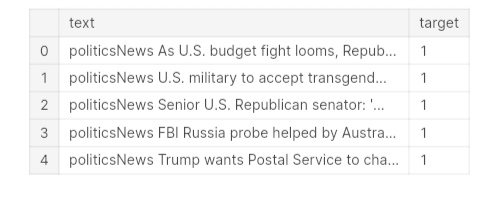
del data['title']

del data['subject']

del data['date']

data.head()

**out[10]:**



**in[11]:**

first\_text = data.text[10]

first\_text

**out[12]:**

**'**politicsNews Jones certified U.S. Senate winner despite Moore challenge (Reuters) - Alabama officials on Thursday certified Democrat Doug Jones the winner of the state’s U.S. Senate race, after a state judge denied a challenge by Republican Roy Moore, whose campaign was derailed by accusations of sexual misconduct with teenage girls. Jones won the vacant seat by about 22,000 votes, or 1.6 percentage points, election officials said. That made him the first Democrat in a quarter of a century to win a Senate seat in Alabama. The seat was previously held by Republican Jeff Sessions, who was tapped by U.S. President Donald Trump as attorney general. A state canvassing board composed of Alabama Secretary of State John Merrill, Governor Kay Ivey and Attorney General Steve Marshall certified the election results. Seating Jones will narrow the Republican majority in the Senate to 51 of 100 seats. In a statement, Jones called his victory “a new chapter” and pledged to work with both parties. Moore declined to concede defeat even after Trump urged him to do so. He stood by claims of a fraudulent election in a statement released after the certification and said he had no regrets, media outlets reported. An Alabama judge denied Moore’s request to block certification of the results of the Dec. 12 election in a decision shortly before the canvassing board met. Moore’s challenge alleged there had been potential voter fraud that denied him a chance of victory. His filing on Wednesday in the Montgomery Circuit Court sought to halt the meeting scheduled to ratify Jones’ win on Thursday. Moore could ask for a recount, in addition to possible other court challenges, Merrill said in an interview with Fox News Channel. He would have to complete paperwork “within a timed period” and show he has the money for a challenge, Merrill said. “We’ve not been notified yet of their intention to do that,” Merrill said. Regarding the claim of voter fraud, Merrill told CNN that more than 100 cases had been reported. “We’ve adjudicated more than 60 of those. We will continue to do that,” he said. Republican lawmakers in Washington had distanced themselves from Moore and called for him to drop out of the race after several women accused him of sexual assault or misconduct dating back to when they were teenagers and he was in his early 30s. Moore has denied wrongdoing and Reuters has not been able to independently verify the allegations.