FAKE NEWS DETECTION USING MACHINE LEARNING

PROJECT REPORT

M.LOGASHRI

FAKE NEWS DETECTION USING MACHINE LEARNING

INTRODUCTION:

In today's digital age, the rapid spread of misinformation through social media and online platforms has made **fake news detection** a critical challenge. Fake news, characterized by false or misleading information presented as authentic, can cause social unrest, manipulate public opinion, and undermine trust in reliable sources. Traditional manual fact-checking methods are inefficient in handling the vast volume of online content. This project leverages **Machine Learning (ML)** and **Natural Language Processing (NLP)** techniques to automatically classify news articles as **real** or **fake**. Using methods such as **TF-IDF Vectorization** for feature extraction and **supervised learning algorithms** like **Logistic Regression**, the system aims to analyze patterns in text data, accurately predict news authenticity, and reduce the spread of misinformation efficiently.

LIBRARIES USED:

1.Pandas

- Purpose: Data manipulation and analysis.
- Usage: Loading the dataset, handling missing values, and combining text columns for better analysis.

2.NumPy

- Purpose: Numerical computations and array operations.

- Usage: Efficient handling of numerical data and mathematical operations on arrays.

3.re (Regular Expressions)

- Purpose: Text preprocessing and pattern matching.
- Usage: Cleaning text data by removing special characters, digits, and unwanted patterns.

4.NLTK (Natural Language Toolkit)

- Purpose: Text processing and natural language understanding.
- Usage: Removing stopwords and performing text normalization.

5.Scikit-learn

- Purpose: Machine learning model training, evaluation, and text vectorization.
- Key Modules Used:
    - TfidfVectorizer: Convert text data into numerical vectors.
    - train_test_split: Split data into training and testing sets.
    - LogisticRegression: Build a classification model.
    - accuracy_score, confusion_matrix, classification_report: Evaluate model performance.

6.Matplotlib & Seaborn

- Purpose: Data visualization and graphical analysis.
- Usage: Plotting confusion matrices and understanding model performance visually.

7.Pickle

- Purpose: Model and object serialization.
- Usage: Save and load trained models and vectorizers for future predictions.

8.Input/Output Modules (Colab-Specific)

- files (from google.colab): Upload datasets directly in Colab.
- Usage: Upload files in the Colab environment for dataset integration.

PROGRAM

Step 1: Import Required Libraries

```
import pandas as pd
import numpy as np
import re
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import matplotlib.pyplot as plt
import seaborn as sns
nltk.download('stopwords')
```

Step 2: Load and Explore Dataset

1.Upload Dataset to Colab

```
from google.colab import files
```

```
uploaded = files.upload()
```

2.Load Dataset into Pandas

```
df = pd.read_csv('news.csv')
print(df.head())
print(df.info())
print(df.isnull().sum())
```

3. Handle Missing Values

```
df = df[['title', 'text']].dropna()
df['content'] = df['title'] + ' ' + df['text']
print(df.head())
```

4. Define Labels (if absent)

```
df['label'] = df['title'].apply(lambda x: 1 if 'fake' in x.lower() else 0)
print(df['label'].value_counts())
```

Step 3: Data Preprocessing

```
def clean_text(text):
    text = re.sub(r'\W', ' ', text)  # Remove special characters
    text = text.lower()  # Lowercase text
    text = re.sub(r'\s+[a-zA-Z]\s+', ' ', text)  # Remove single letters
    text = re.sub(r'\^[a-zA-Z]\s+', ' ', text)  # Remove single letters at the start
    text = re.sub(r'\s+', ' ', text)  # Remove extra spaces
```

```
    return text
df['content'] = df['content'].apply(clean_text)
```

Step 4: Feature Extraction

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(stop_words=stopwords.words('english'), max_features=5000)
X = vectorizer.fit_transform(df['content']).toarray()
y = df['label']
```

Step 5: Train-Test Split

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Step 6: Train the Model

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Step 7: Evaluate the Model

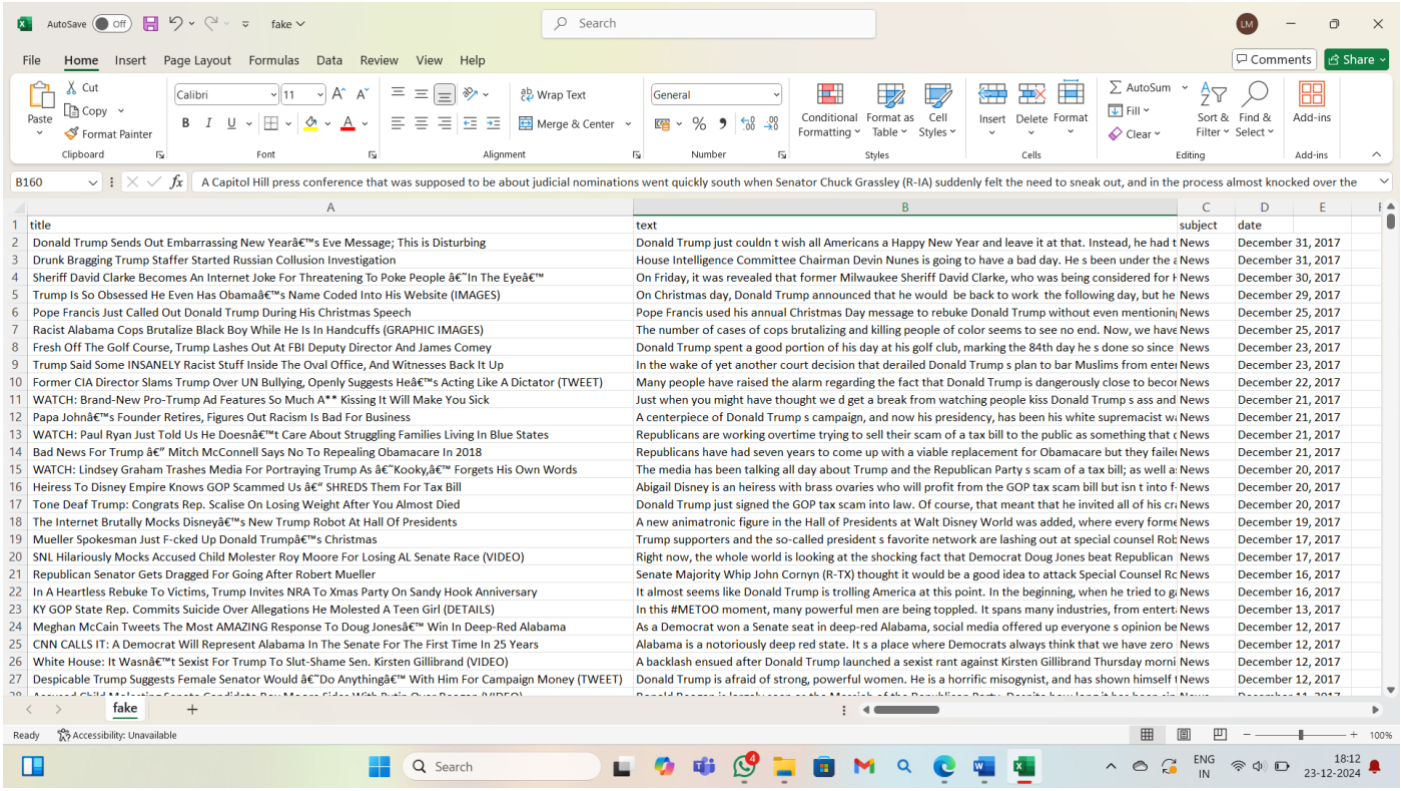
```
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
print("Accuracy:", accuracy_score(y_test, y_pred))
```

```
print("Classification Report:\n", classification_report(y_test, y_pred))
conf_matrix = confusion_matrix(y_test, y_pred)
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues')
plt.title('Confusion Matrix')
plt.show()
```

Step 8: Test with New Data

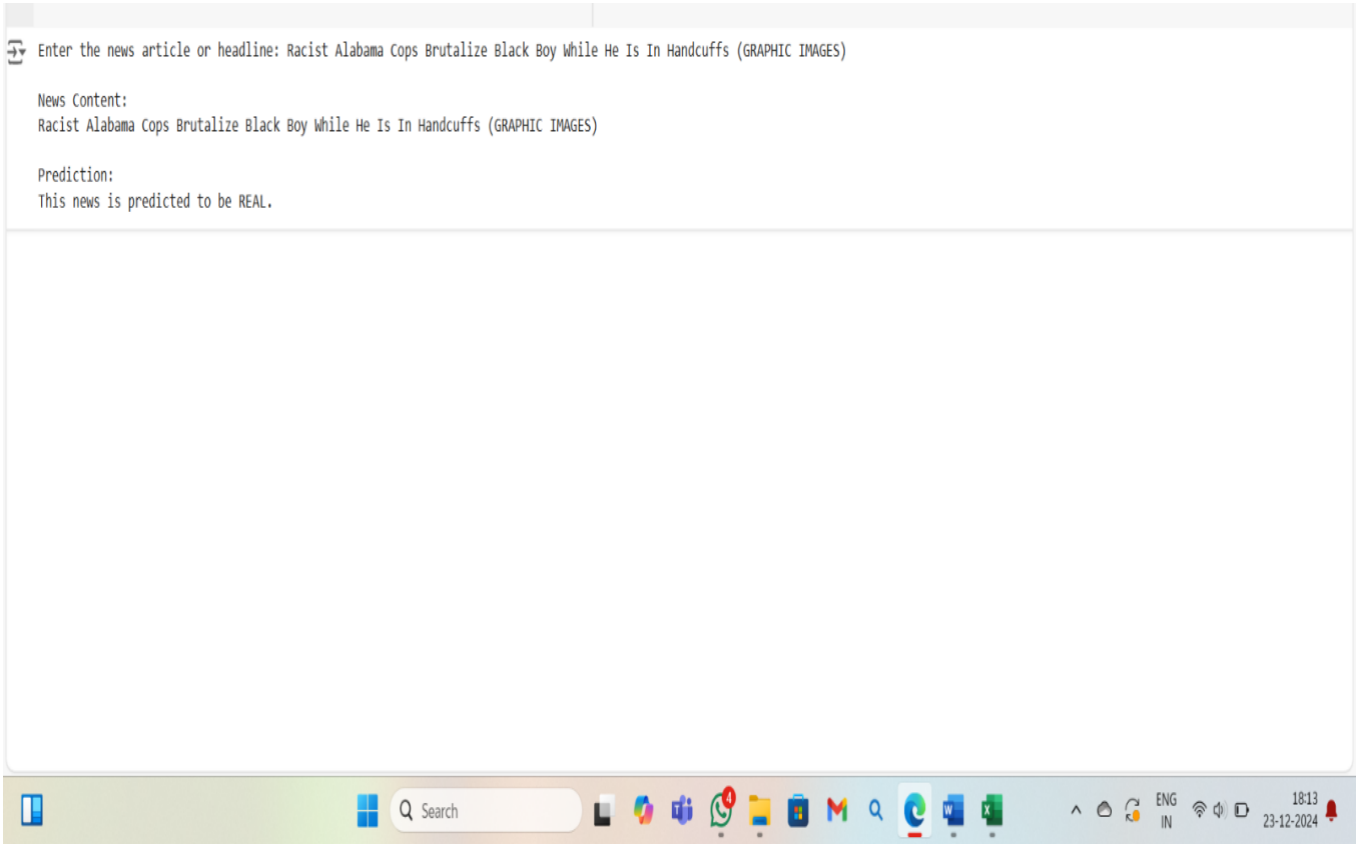
```
user_news = input("Enter the news article or headline: ")
print("\nNews Content:")
print(user_news)
def clean_input_text(text):
    text = re.sub(r'\W', ' ', text)  # Remove special characters
    text = text.lower()  # Convert to lowercase
    text = re.sub(r'\s+[a-zA-Z]\s+', ' ', text)  # Remove single letters
    text = re.sub(r'\^[a-zA-Z]\s+', ' ', text)  # Remove single letters at start
    text = re.sub(r'\s+', ' ', text)  # Remove extra spaces
    return text
cleaned_news = clean_input_text(user_news)
transformed_news = vectorizer.transform([cleaned_news]).toarray()
prediction = model.predict(transformed_news)
print("\nPrediction:")
if prediction[0] == 1:
    print("This news is predicted to be FAKE.")
else:
    print("This news is predicted to be REAL.")
```

Screenshot : news.csv



	A	B	C	D
1	title	text	subject	date
2	Donald Trump Sends Out Embarrassing New Yearâ€™s Eve Message; This is Disturbing	Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had t	News	December 31, 2017
3	Drunk Bragging Trump Staffer Started Russian Collusion Investigation	House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He s been under the a	News	December 31, 2017
4	Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People â€˜In The Eyeâ€™	On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who was being considered for H	News	December 30, 2017
5	Trump Is So Obsessed He Even Has Obamaâ€™s Name Coded Into His Website (IMAGES)	On Christmas day, Donald Trump announced that he would be back to work the following day, but he	News	December 29, 2017
6	Pope Francis Just Called Out Donald Trump During His Christmas Speech	Pope Francis used his annual Christmas Day message to rebuke Donald Trump without even mentioning	News	December 25, 2017
7	Racist Alabama Cops Brutalize Black Boy While He Is In Handcuffs (GRAPHIC IMAGES)	The number of cases of cops brutalizing and killing people of color seems to see no end. Now, we have	News	December 25, 2017
8	Fresh Off The Golf Course, Trump Lashes Out At FBI Deputy Director And James Comey	Donald Trump spent a good portion of his day at his golf club, marking the 84th day he s done so since	News	December 23, 2017
9	Trump Said Some INSANELY Racist Stuff Inside The Oval Office, And Witnesses Back It Up	In the wake of yet another court decision that derailed Donald Trump s plan to bar Muslims from enter	News	December 23, 2017
10	Former CIA Director Slams Trump Over UN Bullying, Openly Suggests Heâ€™s Acting Like A Dictator (TWEET)	Many people have raised the alarm regarding the fact that Donald Trump is dangerously close to becor	News	December 22, 2017
11	WATCH: Brand-New Pro-Trump Ad Features So Much A** Kissing It Will Make You Sick	Just when you might have thought we d get a break from watching people kiss Donald Trump s ass and	News	December 21, 2017
12	Papa Johnâ€™s Founder Retires, Figures Out Racism Is Bad For Business	A centerpiece of Donald Trump s campaign, and now his presidency, has been his white supremacist wi	News	December 21, 2017
13	WATCH: Paul Ryan Just Told Us He Doesnâ€™t Care About Struggling Families Living In Blue States	Republicans are working overtime trying to sell their scam of a tax bill to the public as something that c	News	December 21, 2017
14	Bad News For Trump â€" Mitch McConnell Says No To Repealing Obamacare In 2018	Republicans have had seven years to come up with a viable replacement for Obamacare but they faile	News	December 21, 2017
15	WATCH: Lindsey Graham Trashes Media For Portraying Trump As â€˜Kooky,â€™ Forgets His Own Words	The media has been talking all day about Trump and the Republican Party s scam of a tax bill; as well a	News	December 20, 2017
16	Heiress To Disney Empire Knows GOP Scammed Us â€" SHREDS Them For Tax Bill	Abigail Disney is an heiress with brass ovaries who will profit from the GOP tax scam bill but isn t into f	News	December 20, 2017
17	Tone Deaf Trump: Congrats Rep. Scalise On Losing Weight After You Almost Died	Donald Trump just signed the GOP tax scam into law. Of course, that meant that he invited all of his cr	News	December 20, 2017
18	The Internet Brutally Mocks Disneyâ€™s New Trump Robot At Hall Of Presidents	A new animatronic figure in the Hall of Presidents at Walt Disney World was added, where every forme	News	December 19, 2017
19	Mueller Spokesman Just F-cked Up Donald Trumpâ€™s Christmas	Trump supporters and the so-called president s favorite network are lashing out at special counsel Rob	News	December 17, 2017
20	SNL Hilariously Mocks Accused Child Molester Roy Moore For Losing AL Senate Race (VIDEO)	Right now, the whole world is looking at the shocking fact that Democrat Doug Jones beat Republican	News	December 17, 2017
21	Republican Senator Gets Dragged For Going After Robert Mueller	Senate Majority Whip John Cornyn (R-TX) thought it would be a good idea to attack Special Counsel Rc	News	December 16, 2017
22	In A Heartless Rebuke To Victims, Trump Invites NRA To Xmas Party On Sandy Hook Anniversary	It almost seems like Donald Trump is trolling America at this point. In the beginning, when he tried to g	News	December 16, 2017
23	KY GOP State Rep. Commits Suicide Over Allegations He Molested A Teen Girl (DETAILS)	In this #METOO moment, many powerful men are being toppled. It spans many industries, from entert	News	December 13, 2017
24	Meghan McCain Tweets The Most AMAZING Response To Doug Jonesâ€™ Win In Deep-Red Alabama	As a Democrat won a Senate seat in deep-red Alabama, social media offered up everyone s opinion be	News	December 12, 2017
25	CNN CALLS IT: A Democrat Will Represent Alabama In The Senate For The First Time In 25 Years	Alabama is a notoriously deep red state. It s a place where Democrats always think that we have zero	News	December 12, 2017
26	White House: It Wasnâ€™t Sexist For Trump To Slut-Shame Sen. Kirsten Gillibrand (VIDEO)	A backlash ensued after Donald Trump launched a sexist rant against Kirsten Gillibrand Thursday morni	News	December 12, 2017
27	Despicable Trump Suggests Female Senator Would â€˜Do Anythingâ€™ With Him For Campaign Money (TWEET)	Donald Trump is afraid of strong, powerful women. He is a horrific misogynist, and has shown himself t	News	December 12, 2017

Output screenshot:



```
Enter the news article or headline: Racist Alabama Cops Brutalize Black Boy While He Is In Handcuffs (GRAPHIC IMAGES)

News Content:
Racist Alabama Cops Brutalize Black Boy While He Is In Handcuffs (GRAPHIC IMAGES)

Prediction:
This news is predicted to be REAL.
```

Conclusion:

The Fake News Detection project showcases how Machine Learning and Natural Language Processing can help address the problem of misinformation online. By using models like Logistic Regression and Naive Bayes, and employing TF-IDF Vectorization, the system can efficiently classify news articles as real or fake. This automated approach reduces the reliance on manual fact-checking. While the model performs well, further improvements can be made by exploring more advanced algorithms. Ultimately, this project contributes to the fight against fake news by providing a reliable tool for verifying information. It demonstrates the potential of technology in ensuring accurate, trustworthy news consumption.