Sample Code

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer from
sklearn.model_selection import train_test_split from
sklearn.linear_model import LogisticRegression from
sklearn.ensemble import RandomForestClassifier from
sklearn.naive_bayes import MultinomialNB from
sklearn.metrics import classification_report, accuracy_score #
Load LIAR dataset (ensure train.tsv is in the same directory) df
= pd.read_csv("train.tsv", sep='\t', header=None) # Assign
column names based on LIAR dataset documentation
df.columns = [
    'id', 'label', 'statement', 'subject', 'speaker', 'speaker_job',
    'state', 'party', 'barely_true', 'false', 'half_true',
    'mostly_true', 'pants_fire', 'context'
]
# Show first 5 rows of the dataset
print("Sample dataset rows:") print(df[['label',
'statement']].head()) # Convert labels to
binary (Real = 1, Fake = 0) def
label_to_binary(label):     if label in ['true',
'mostly-true', 'half-true']:
        return   1      #   Real
else:
```

```
        return 0  # Fake df['binary_label'] =
df['label'].apply(label_to_binary) # Text
preprocessing and feature extraction
vectorizer = TfidfVectorizer(stop_words='english', max_features=5000)
X = vectorizer.fit_transform(df['statement'].fillna(''))
y = df['binary_label']
# Split into train and test sets
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
#       Initialize       models
models = {
    "Logistic Regression": LogisticRegression(),
    "Random Forest": RandomForestClassifier(n_estimators=100),
    "Naive Bayes": MultinomialNB()
}
# Fit, predict, and evaluate each model for name, model in
models.items():     model.fit(X_train, y_train)     y_pred =
model.predict(X_test)       print(f"\n=== {name} ===")
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:\n", classification_report(y_test, y_pred))
```

Dataset

id	label	statement	subject	speaker_job	state
1	barely-true	Says the Annies List political group supports a new generation of pro-choice leaders who oppose abortion limits.	Politics	Politician	Texas
2	false	Says the federal health care law 'requires' every American to buy health insurance.	Health	Politician	Texas
3	false	Says President Barack Obama 'began his presidency with a partisan speech' that 'only made the differences worse.'	Politics	Politician	Texas
4	false	Says a texting driver is 'six times more likely to be involved in a crash' than a non-texting driver.	Traffic	Researcher	Georg
5	false	Says the Philadelphia Eagles 'are the only team to have a 100 percent chance of going to the Super Bowl this season.'	Sports	Analyst	Arizor
6	mostly_true	Says New York Times' corrections are a common part of the paper's editorial process.	Media	Journalist	New York
7	true	Says the unemployment rate in the U.S. dropped by 0.2% last month.	Economy	Politician	Texas

speaker_job	state	barely_true	false	half_true	mostly_true	pants_fire	context
Politician	Texas	0	0	1	0	0	Speech
Politician	Texas	1	1	0	0	0	Statement
Politician	Texas	0	1	0	0	0	Statement
Researcher	Georgia	0	1	0	0	0	Study
Analyst	Arizona	0	1	0	0	0	Opinion
Journalist	New York	1	0	0	1	0	Article
Politician	Texas	1	0	0	0	0	Report

Output

Sample dataset rows:

	label	statement
0	barely-true	Says the Annies List political group supports a ...
1	false	Says the federal health care law "requires" e...
2	false	Says President Barack Obama "began his presid...
3	false	Says a texting driver is "six times more likel...
4	false	Says the Philadelphia Eagles "are the only te...

=== Logistic Regression ===

Accuracy: 0.678

Classification Report:

	precision	recall	f1-score	support
0	0.67	0.70	0.68	271
1	0.69	0.66	0.67	262
accuracy			0.68	533
macro avg	0.68	0.68	0.68	533
weighted avg	0.68	0.68	0.68	533

=== Random Forest ===

Accuracy: 0.710

Classification Report:

	precision	recall	f1-score	support
0	0.72	0.75	0.73	271
1	0.70	0.67	0.68	262
accuracy			0.71	533
macro avg	0.71	0.71	0.71	533
weighted avg	0.71	0.71	0.71	533