

# nlp-8

January 5, 2026

## 0.0.1 Assignment N0. 2

```
[1]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

```
[2]: df=pd.read_csv("Downloads/train.csv")
df.head()
```

```
[2]:      Class Index          Title \
0            3  Wall St. Bears Claw Back Into the Black (Reuters)
1            3  Carlyle Looks Toward Commercial Aerospace (Reu...
2            3  Oil and Economy Cloud Stocks' Outlook (Reuters)
3            3  Iraq Halts Oil Exports from Main Southern Pipe...
4            3  Oil prices soar to all-time record, posing new...
```

```
                  Description
0  Reuters - Short-sellers, Wall Street's dwindle...
1  Reuters - Private investment firm Carlyle Grou...
2  Reuters - Soaring crude prices plus worries\ab...
3  Reuters - Authorities have halted oil export\f...
4  AFP - Tearaway world oil prices, toppling reco...
```

```
[3]: df['Full_Text'] = df['Title'] + " " + df['Description']
```

```
[4]: df
```

```
[4]:      Class Index          Title \
0            3  Wall St. Bears Claw Back Into the Black (Reuters)
1            3  Carlyle Looks Toward Commercial Aerospace (Reu...
2            3  Oil and Economy Cloud Stocks' Outlook (Reuters)
3            3  Iraq Halts Oil Exports from Main Southern Pipe...
4            3  Oil prices soar to all-time record, posing new...
...
...           ...           ...
```

119995	1	Pakistan's Musharraf Says Won't Quit as Army C...
119996	2	Renteria signing a top-shelf deal
119997	2	Saban not going to Dolphins yet
119998	2	Today's NFL games
119999	2	Nets get Carter from Raptors

		Description \
0	Reuters - Short-sellers, Wall Street's dwindli...	
1	Reuters - Private investment firm Carlyle Grou...	
2	Reuters - Soaring crude prices plus worries\ab...	
3	Reuters - Authorities have halted oil export\f...	
4	AFP - Tearaway world oil prices, toppling reco...	

...	...
119995	KARACHI (Reuters) - Pakistani President Perve...
119996	Red Sox general manager Theo Epstein acknowled...
119997	The Miami Dolphins will put their courtship of...
119998	PITTSBURGH at NY GIANTS Time: 1:30 p.m. Line: ...
119999	INDIANAPOLIS -- All-Star Vince Carter was trad...

	Full_Text
0	Wall St. Bears Claw Back Into the Black (Reute...
1	Carlyle Looks Toward Commercial Aerospace (Reu...
2	Oil and Economy Cloud Stocks' Outlook (Reuters...
3	Iraq Halts Oil Exports from Main Southern Pipe...
4	Oil prices soar to all-time record, posing new...

  

...	...
119995	Pakistan's Musharraf Says Won't Quit as Army C...
119996	Renteria signing a top-shelf deal Red Sox gene...
119997	Saban not going to Dolphins yet The Miami Dolp...
119998	Today's NFL games PITTSBURGH at NY GIANTS Time...
119999	Nets get Carter from Raptors INDIANAPOLIS -- A...

[120000 rows x 4 columns]

```
[5]: tfidf = TfidfVectorizer(stop_words='english', max_features=5000)
```

```
[6]: X = tfidf.fit_transform(df['Full_Text'])
```

```
[7]: y = df['Class Index']
```

```
[8]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

```
[9]: model = MultinomialNB()
```

```
[10]: model.fit(X_train, y_train)
```

```
[10]: MultinomialNB()
```

```
[11]: y_pred = model.predict(X_test)
      print(accuracy_score(y_test, y_pred))
```

0.895875

```
[12]: model = RandomForestClassifier()
```

```
[13]: model.fit(X_train, y_train)
```

```
[13]: RandomForestClassifier()
```

```
[14]: y_pred = model.predict(X_test)
      print(accuracy_score(y_test, y_pred))
```

0.881875

```
[15]: model = LogisticRegression(max_iter=1000)
```

```
[16]: model.fit(X_train, y_train)
```

```
[16]: LogisticRegression(max_iter=1000)
```

```
[17]: y_pred = model.predict(X_test)
```

```
[18]: print(accuracy_score(y_test, y_pred))
```

0.9094166666666667

```
[ ]:
```