

## ▼ Fake News Detection

### ▼ Task 2

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from colorama import Fore, Style
```

```
data = pd.read_csv('train_data_mal_fake_detect.csv')
```

```
# Fill missing values with an empty string
data['News'] = data['News'].fillna('')
```

```
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(data['News'])
y = data['Label']
```

```
X_train, X_dev, y_train, y_dev = train_test_split(X, y, test_size=0.2, random_state=42)
```

### ▼ Logistic Regression

```
param_grid_lr = {'C': [0.001, 0.01, 0.1, 1, 10, 100],
                  'solver': ['liblinear', 'lbfgs', 'newton-cg', 'sag', 'saga']}
```

```
lr_classifier = LogisticRegression(max_iter=1000, random_state=42)
grid_search_lr = GridSearchCV(lr_classifier, param_grid_lr, cv=5, scoring='accuracy', n_jobs=-1)
grid_search_lr.fit(X_train, y_train)
```

C:\Python310\lib\site-packages\sklearn\model\_selection\\_split.py:700: UserWarning: The least populated class in y has only 1 members: warnings.warn(

```
GridSearchCV
  estimator: LogisticRegression
    LogisticRegression
```

```
# Predict on the development set
best_classifier_lr = grid_search_lr.best_estimator_
lr_predictions = best_classifier_lr.predict(X_dev)
lr_accuracy = accuracy_score(y_dev, lr_predictions)
```

### ▼ Random Forest

```
param_grid_rf = {'n_estimators': [50, 100, 200],
                  'max_depth': [None, 10, 20, 30],
                  'min_samples_split': [2, 5, 10],
                  'min_samples_leaf': [1, 2, 4]}
```

```
rf_classifier = RandomForestClassifier(random_state=42)
grid_search = GridSearchCV(rf_classifier, param_grid_rf, cv=5, scoring='accuracy', n_jobs=-1)
grid_search.fit(X_train, y_train)
```

C:\Python310\lib\site-packages\sklearn\model\_selection\\_split.py:700: UserWarning: The least populated class in y has only 1 members: warnings.warn(

```
GridSearchCV
  estimator: RandomForestClassifier
    RandomForestClassifier
```

```
# Predict on the development set
best_classifier_rf = grid_search.best_estimator_
rf_predictions = best_classifier_rf.predict(X_dev)
rf_accuracy = accuracy_score(y_dev, rf_predictions)
```

▼ SVM

```
svm_classifier = SVC(kernel='rbf', random_state=42)
svm_classifier.fit(X_train, y_train)
svm_predictions = svm_classifier.predict(X_dev)
svm_accuracy = accuracy_score(y_dev, svm_predictions)
```

▼ Naive Bayes

```
nb_classifier = MultinomialNB()
nb_classifier.fit(X_train, y_train)
nb_predictions = nb_classifier.predict(X_dev)
nb_accuracy = accuracy_score(y_dev, nb_predictions)
```

▼ Result and Inference

```
print("Logistic Regression Accuracy:", lr_accuracy)
print("Random Forest Accuracy:", rf_accuracy)
print("Support Vector Machine Accuracy:", svm_accuracy)
print("Naive Bayes Accuracy:", nb_accuracy)
```

Logistic Regression Accuracy: 0.655688622754491  
Random Forest Accuracy: 0.655688622754491  
Support Vector Machine Accuracy: 0.655688622754491  
Naive Bayes Accuracy: 0.6526946107784432

```
test_without_label = pd.read_csv('Fake_test_without_labels.csv')
```

```
x = vectorizer.transform(test_without_label['text'])
```

```
predictions_without_label_nb = classifier_nb.predict(x)
```

```
results_without_label_nb = pd.DataFrame({
    'Text': test_without_label['text'],
    'Predicted': predictions_without_label_nb
})
```

```
results_without_label_nb
```

	Text	Predicted
0	5000 ഉള്ള പോൾ ലോഗ്ഡ്വൻ ഇപ്പോള് 250000 എന്താ...	Fake
1	ഓഷോ രജനീഷ് പറഞ്ഞപോലെ എനിക്കപ്പോൾ തോന്നിയത് അ...	original
2	ചേട്ടാ വാർത്ത വയ്ക്കുന്നത് കേരളത്തിലാണ് സം...	Fake
3	Shame for entire Woman&#39	Fake
4	135 code janaghal andhu wide business cheythal...	Fake
...	...	...
1014	Correct ...China cheitha ...weapon spread ing ...	Fake
1015	ഈ WHO പറയുന്നതനുസരിച്ചു ചികിത്സയും ലോക്ക് ഡൗൺ ...	Fake
1016	Mask illa aarkum 😞😞	original
1017	ഇയാളെ കൊറോണ. രോഗി കൾ കിടയിൽ. ഇടാമായിരുന്നു---!!	Fake
1018	Kulathrikalkku badilayiAlavilathikal ...	original

1019 rows × 2 columns

▼ Applying SVM in test data which is without label.

```
test_without_label = pd.read_csv('test-data.csv')
```

```
test_without_label['News'] = test_without_label['News'].fillna('')
```

```
x = vectorizer.transform(test_without_label['News'])
```

```
predictions_without_label_nb = svm_classifier.predict(x)
```

```
results_without_label_nb = pd.DataFrame({
    'News': test_without_label['News'],
    'Predicted': predictions_without_label_nb
})
```

```
results_without_label_nb
```

	News	Predicted
0	ചന്ദനക്കുറിയണിഞ്ഞ് വിഎസ് അച്യുതാനന്ദൻ.	FALSE
1	നെറ്റിയിൽ ചന്ദനക്കുറിയിട്ട് വിഎസ് അച്യുതാനന്...	FALSE
2	മുഖ്യമന്ത്രി പിണറായി വിജയനെതിരെ വിഎസ് അച്യുതാന...	FALSE
3	വിശ്വസിച്ച് കൂടെ നിന്ന സഖാവ് അലന്റേയും സഖാവ് താഹ...	FALSE
4	പിണറായി വിജയനെപ്പറ്റി വിഎസ് അച്യുതാനന്ദൻ തരംതാ...	FALSE
...	...	...
245	തിരുവനന്തപുരം, കോവളത്ത് നൂറുകണക്കിന് സിപിഎമ്മു...	FALSE
246	പാലക്കാട് ഇരുചക്രവാഹനങ്ങളിൽ പുരുഷന്മാരുടെ പിൻസ...	FALSE
247	ബിവറേജസ് ഔട്ട്ലെറ്റുകൾ തുറന്ന ആദ്യ ദിനത്തില...	FALSE
248	ഇക്കഴിഞ്ഞ ദിവസം വയനാട് സന്ദർശിക്കാനെത്തിയ രാ...	FALSE
249	ഇഡി ചോദ്യം ചെയ്യലിനു ശേഷം വയനാട്ടിലെത്തിയ രാഹു...	FALSE

250 rows × 2 columns

```
results_without_label_nb.to_csv('output_file_task2.csv', index=False)
```