In [*]:

```
from sklearn.datasets import fetch_20newsgroups
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_selection import SelectKBest, chi2, mutual_info_classi
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import make_pipeline
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# Load 20NG dataset
newsgroups = fetch_20newsgroups(subset='all')
X, y = newsgroups.data, newsgroups.target

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ra

# Preprocess using TF-IDF vectorization
vectorizer = TfidfVectorizer(stop_words='english', max_df=0.5)

# Define classifiers
clf = MultinomialNB()

# For chi2 criteria
chi2_selector = SelectKBest(chi2, k=200)
chi2_pipeline = make_pipeline(vectorizer, chi2_selector, clf)
chi2_pipeline.fit(X_train, y_train)
chi2_pred = chi2_pipeline.predict(X_test)
chi2_acc = accuracy_score(y_test, chi2_pred)
print(f"Accuracy with chi2 criteria: {chi2_acc:.4f}")

# For mutual information criteria
mi_selector = SelectKBest(mutual_info_classif, k=200)
mi_pipeline = make_pipeline(vectorizer, mi_selector, clf)
mi_pipeline.fit(X_train, y_train)
mi_pred = mi_pipeline.predict(X_test)
mi_acc = accuracy_score(y_test, mi_pred)
print(f"Accuracy with mutual information criteria: {mi_acc:.4f}")

# Note: You should have a previously recorded accuracy from HW3A-PB1 to com
```

```
Accuracy with chi2 criteria: 0.6045
```