СРС 5 Машиналық оқытуды қолданып мәтінді өңдеуді автоматтандыру

!pip install numpy pandas scikit-learn nltk

import nltk

from sklearn.feature\_extraction.text import CountVectorizer

from nltk.corpus import stopwords

import string

nltk.download('stopwords')

def preprocess\_text(text):

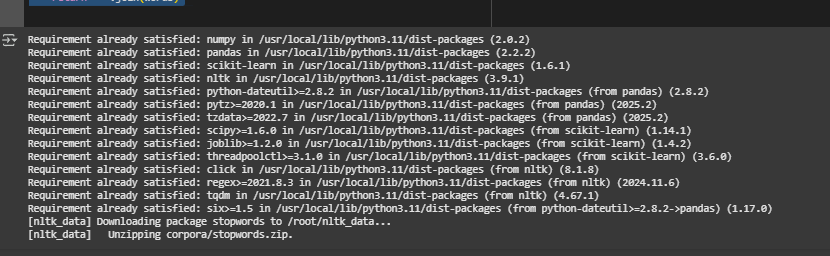
    text = text.lower()  # Барлық әріпті кіші әріпке ауыстыру

    text = ''.join([char for char in text if char not in string.punctuation])  # Тыныс белгілерін алып тастау

    words = text.split()

    words = [word for word in words if word not in stopwords.words('english')]  # Stop сөздерді жою

    return ' '.join(words)



texts = [

    "I love machine learning!",

    "Text processing is fun.",

    "Python is a powerful language for data science.",

    "Automating text tasks is very useful."

]

# Мәтінді алдын ала өңдеу

cleaned\_texts = [preprocess\_text(text) for text in texts]

# Мәтінді векторизациялау

vectorizer = CountVectorizer()

X = vectorizer.fit\_transform(cleaned\_texts)

from sklearn.naive\_bayes import MultinomialNB

from sklearn.model\_selection import train\_test\_split

# Жалған белгілер (мысал үшін)

y = [1, 0, 1, 0]

# Мәліметтерді бөлу

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2)

# Модельді оқыту

model = MultinomialNB()

model.fit(X\_train, y\_train)

# Болжау

predictions = model.predict(X\_test)

print("Predictions:", predictions)

from sklearn.pipeline import Pipeline

text\_clf = Pipeline([

    ('vect', CountVectorizer()),

    ('clf', MultinomialNB())

])

# Модельді оқыту

text\_clf.fit(cleaned\_texts, y)

# Жаңа мәтінмен болжау

new\_text = ["Machine learning automates tasks."]

processed = [preprocess\_text(t) for t in new\_text]

print(text\_clf.predict(processed))

print("Predicted:", predictions)

print("Actual:   ", y\_test)

from sklearn.metrics import accuracy\_score

accuracy = accuracy\_score(y\_test, predictions)

print("Accuracy:", accuracy)

from sklearn.metrics import classification\_report

print(classification\_report(y\_test, predictions))

