MACHINE LEARNING TASK 1:

MOVIE GENRE

CLASSIFICATION

Create a machine learning model that can predict the genre of a

movie based on its plot summary or other textual information. You

can use techniques like TF-IDF or word embeddings with classifiers

such as Naive Bayes, Logistic Regression, or Support Vector

Machines.

import pandas as pd

import re

from nltk.corpus import stopwords

from nltk.stem import PorterStemmer

from nltk.tokenize import word\_tokenize

# Load dataset

df = pd.read\_csv('movies.csv')

# Initialize stopwords and stemmer

stop\_words = set(stopwords.words('english'))

stemmer = PorterStemmer()

def preprocess\_text(text):

# Remove special characters and lowercase

text = re.sub(r'\W', ' ', text)

text = text.lower()

# Tokenization

tokens = word\_tokenize(text)

# Remove stopwords and stem words

tokens = [stemmer.stem(word) for word in tokens if word not in stop\_words]

return ' '.join(tokens)

# Apply preprocessing

df['processed\_plot'] = df['plot\_summary'].apply(preprocess\_text)

Example:

from sklearn.svm import SVC

# Initialize and train the classifier

svc\_classifier = SVC(kernel='linear')

svc\_classifier.fit(X\_train, y\_train)

# Predict and evaluate

y\_pred = svc\_classifier.predict(X\_test)

print("Accuracy:", accuracy\_score(y\_test, y\_pred))

print("Classification Report:\n", classification\_report(y\_test, y\_pred))