# train\_model.py

import pandas as pd

import pickle

from sklearn.pipeline import Pipeline

from sklearn.compose import ColumnTransformer

from sklearn.preprocessing import OneHotEncoder

from xgboost import XGBRegressor

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

# Load cleaned and pivoted dataset

df = pd.read\_csv("Crop\_Production\_Cleaned.csv")

pivot\_df = df.pivot\_table(index=['Area', 'Item', 'Year'], columns='Element', values='Value').reset\_index()

pivot\_df.dropna(subset=['Production', 'Yield', 'Area harvested'], inplace=True)

X = pivot\_df[['Area', 'Item', 'Year', 'Yield', 'Area harvested']]

y = pivot\_df['Production']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

categorical = ['Area', 'Item']

numerical = ['Year', 'Yield', 'Area harvested']

preprocessor = ColumnTransformer([

    ("cat", OneHotEncoder(handle\_unknown='ignore'), categorical)

], remainder='passthrough')

pipeline = Pipeline([

    ("preprocessor", preprocessor),

    ("model", XGBRegressor(n\_estimators=100, learning\_rate=0.1, random\_state=42))

])

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

with open("final\_model.pkl", "wb") as f:

    pickle.dump(pipeline, f)

print("✅ Model trained and saved as final\_model.pkl")