# train\_model.py

import pandas as pd

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

from sklearn.ensemble import RandomForestClassifier

from sklearn.preprocessing import LabelEncoder

import joblib

# Load dataset

data = pd.read\_csv("student\_data.csv") # Replace with your dataset path

# Basic preprocessing

data = data.dropna()

label\_enc = LabelEncoder()

# Encode categorical columns

categorical\_cols = data.select\_dtypes(include=['object']).columns

for col in categorical\_cols:

data[col] = label\_enc.fit\_transform(data[col])

# Feature selection

X = data.drop('G3', axis=1) # Assuming 'G3' is the target

y = data['G3']

y = y.apply(lambda grade: 1 if grade >= 10 else 0) # Binary classification

# Train-test split

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

# Model training

model = RandomForestClassifier()

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

# Save model

joblib.dump(model, 'student\_model.pkl')