# 极简线性回归模板（自动过滤无用特征）

# 功能：1.删除无用列 2.训练模型 3.输出submit.csv

import pandas as pd

from sklearn.linear\_model import LinearRegression

# 1. 加载数据

train\_data = pd.read\_csv("/kaggle/input/houkongtest/stock\_train\_data.csv") # 训练集

test\_data = pd.read\_csv("/kaggle/input/houkongtest/stock\_test\_data.csv") # 测试集

# 2. 指定列用途（根据实际数据修改！）

target\_col = "close" # 替换为你的目标列名

useless\_cols = ["id", "date"] # 指定要删除的无用列名（如ID、日期等）

used\_features = ["open", "high", "low","volume","adj\_close"] # 指定要使用的特征列名

# 3. 数据清洗（删除无用列）

# 训练集处理

X\_train = train\_data[used\_features].values # 只保留有用特征

y\_train = train\_data[target\_col].values # 提取目标列

# 测试集处理（保留ID用于提交，但训练时不使用）

test\_ids = test\_data["date"].values # 假设测试集有id列

X\_test = test\_data[used\_features].values # 使用与训练集相同的特征

# 4. 训练模型

model = LinearRegression()

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

# 5. 预测并保存结果

pd.DataFrame({

"id": test\_ids,

"prediction": model.predict(X\_test)

}).to\_csv("submit.csv", index=False)

print("结果已保存至 submit.csv")

print("使用特征:", used\_features)

print("删除列:", useless\_cols)