学生实验实习报告册



学年学期： 2020-2021 学年(春)学期

课程名称： 数据工程基础实践

学生学院： 计算机学院/人工智能学院

专业班级： 数据科学与大数据技术

学生学号： XXXXXX

学生姓名： 周继聪

联系电话： 18982651030

重庆邮电大学教务处印制

目 录



教师评阅记录表

实验报告

教师评阅记录表



【重要说明】

学生提交报告册最终版时，必须包含此页，否则不予成绩评定。

本报告册模板内容格式除确实因为填写内容改变了布局外，不得变更其余部分的格式，

否则不予成绩评定。

报告是否符合考核规范 ☑ 符合 □ 不符合

报告格式是否符合标准 ☑ 符合 □ 不符合

报告是否完成要求内容 ☑ 是 □ 否

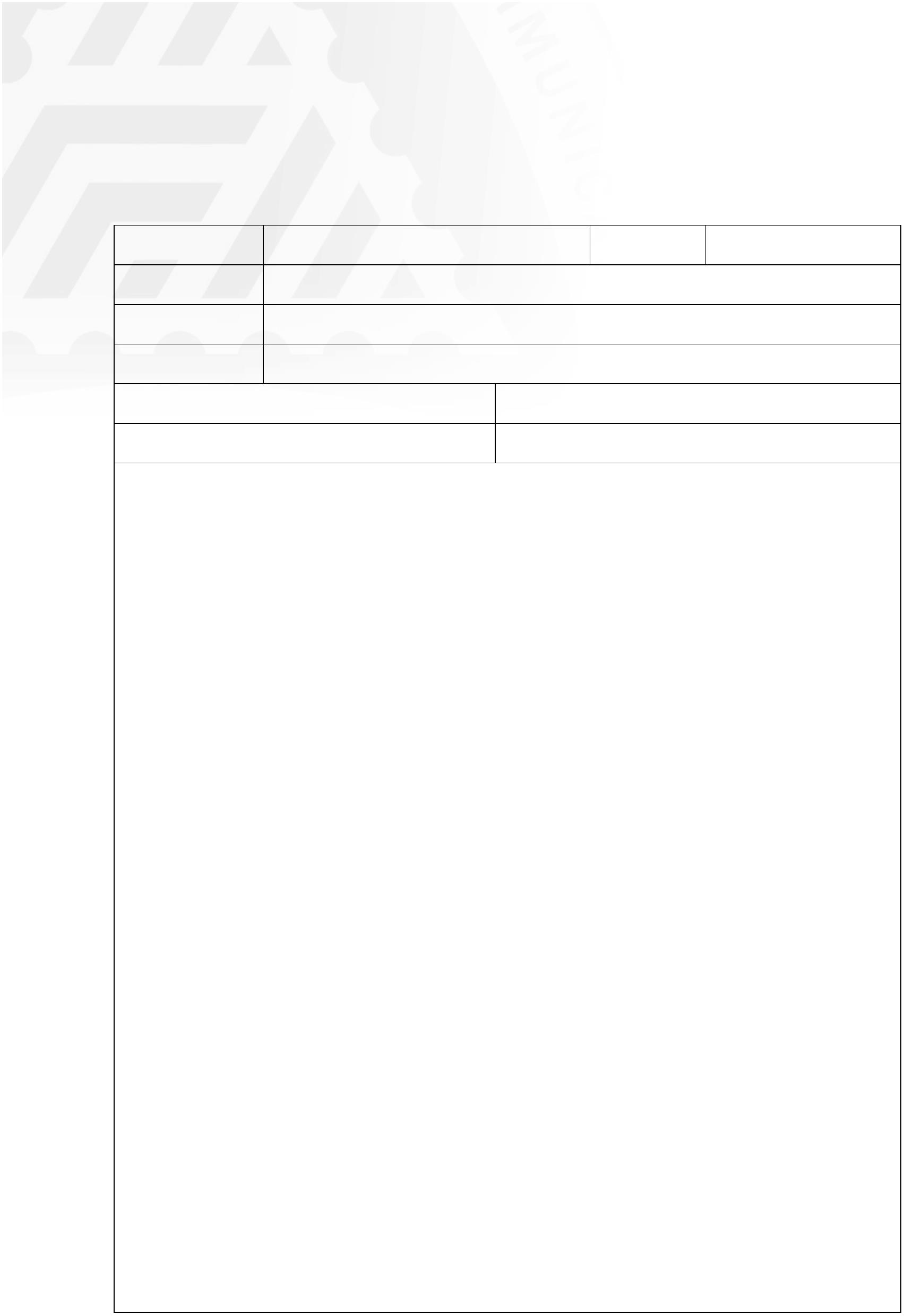
报告评语：

报告成绩：

评阅人签名（签章）

2021 年 7 月 8 日

实验或实习报告



课程名称 数据工程基础实践 课程编号 A2040850

开课学院 计算机学院/人工智能学院

指导教师 王进

实验实习地点 综合实验大楼 B517

学号 姓名

\*\*\*\*\* 周继聪

实验一

一、实验题目及内容

课后作业（二）1：用 Python 实现 F1-score，并自己构建一个简易数据集进行测试

二、实验过程步骤（注意是主要关键步骤，适当文字+截图说明）、实验结果及分析

（一）代码：

# F1-score 实现

import numpy as np

# 构建的简易数据集：

# 实际数据的标签（其中 0 表示反例，1 表示正例）

date\_true = np.random.randint(0, 2, 5000)

date\_true = date\_true.tolist()

# 预测数据的标签

date\_pre = np.random.randint(0, 2, 5000)

date\_pre = date\_pre.tolist()

print("实际标签：", date\_true)

print("预测标签：", date\_pre)

TP = 0

FP = 0

TN = 0

FN = 0

# F1-score 函数

def f1(date\_t, date\_p, mode):

lens = len(date\_pre)

for index in range(lens):

if date\_t[index] == 1:

if date\_t[index] == date\_p[index]:

global TP

TP += 1 # 真正例

else:

global FN



FN += 1 # 假反例

else:

if date\_t[index] == date\_p[index]:

global TN

TN += 1 # 真反例

else:

global FP

FP += 1 # 假正例

if mode == 1: # 求正例 F1\_score

precision\_p = TP / (TP + FP) # 正例查准率

recall\_p = TP / (TP + FN) # 正例查全率

f1\_score\_p = round(2 \* precision\_p \* recall\_p / (recall\_p + precision\_p), 3)

return f1\_score\_p

elif mode == 0: # 求反例 F1\_score

precision\_n = TN / (TN + FN) # 反例查准率

recall\_n = TN / (TN + FP) # 反例查全率

f1\_score\_n = round(2 \* precision\_n \* recall\_n / (recall\_n + precision\_n), 3)

return f1\_score\_n

print("正例 F1-score 为：")

print(f1(date\_true, date\_pre, 1))

print("反例 F1-score 为：")

print(f1(date\_true, date\_pre, 0))

（二）主要步骤：

1.F1-score 主要是通过求得预测数据集的查准率和查全率来计算的，而要想得查准率和查全

率，就要得到预测数据集的真正例 TP，真反例 TN，假正例 FP，假反例 FN 的个数。所以要实现 F1-score，

主要就是求得这四个值。

2.首先要构建一个简易的数据集

date\_true = np.random.randint(0, 2, 5000)

date\_true = date\_true.tolist()

date\_pre = np.random.randint(0, 2, 5000)

date\_pre = date\_pre.tolist()

这里我使用的是随机出 5000 个 0 或者 1 的标签 0 表示反例，1 表示正例。其中 date\_true 是真实

的数据分类，而 date\_pre 是预测的数据分类。然后再把两个数组转化成 list 方便处理。

3.对于这个数据集的思路就是，如果对于一个实际是正例的数据，预测出来同样也是正例，则真

**迅捷PDF转换器**