**综合项目实践课程目标：**

整体目标：完成我国三大城市群（粤港澳大湾区、长三角城市群和京津冀城市群）暴雨内涝事件网页数据的收集、数据预处理、数据分析、模型评价和结果可视化。

算法技能目标：能够应用机器学习、统计分析的相关算法。

编程技能目标：能够使用python语言进行数据的处理、分析和建模；能够使用html和java script进行可视化。

思政目标：深刻认识我国城市暴雨内涝灾害现状，建立防灾意识。

**任务流程：**

一、**网页数据收集：**

1、通过查阅文献了解城市内涝及其危害。

2、了解粤港澳大湾区、长三角城市群和京津冀城市群的地理范围。

3、利用python编写网络爬虫程序，从互联网上搜索和下载暴雨内涝事件的html数据。

二、**网页数据筛选和结构化处理**

1、数据自动筛选：编写python程序实现对网页数据的筛检，删除那些与城市暴雨内涝事件不相关的html文件。可以看作是一个自然语言处理的文本二分类问题。

参考网址：

https://zhuanlan.zhihu.com/p/30945590

https://blog.csdn.net/weixin\_40662229/article/details/80802325

https://blog.csdn.net/qq\_37295506/article/details/86152428

https://blog.csdn.net/weixin\_40240670/article/details/80612813

2、数据结构化处理：针对每个html文件，尽可能地从文件中提取暴雨内涝相关信息（可以通过手动提取）：

（1）暴雨内涝时间

（2）暴雨内涝城市

（3）降雨信息，例如降雨量，降雨强度，降雨历时等

（4）内涝信息，例如内涝严重程度、积水深度、积水点位置、淹没范围，内涝持续时间等

（5）内涝危害信息，例如人员伤亡、受困人员、受灾人员、车辆损失、房屋损失、直接经济损失、交通影响（道路、地铁、列车、机场等）、电力等基础设施影响等）。

将上述暴雨内涝信息保存到excel或csv文件当中。

**三、数据分析**

1、通过编写python程序分析出城市群和各个城市的暴雨内涝事件集中发生的月份和季节。

2、通过编写python程序分析出城市群和各个城市的暴雨内涝事件集中发生的时间段及其与交通早晚高峰时间的重合度。

3、通过编写python程序统计出城市群当中哪些城市更容易发生暴雨内涝。

4、通过编程python程序统计出城市群常见的暴雨内涝危害；根据暴雨内涝危害种类对城市进行排序；分析城市群内涝危害随年份的趋势变化。

**四、暴雨内涝灾害指数评价**

1、构建暴雨内涝灾害指数评价指标体系。

2、利用层次分析法确定指标权重，要求通过python编程实现。

3、确定每个评价指标的打分方法和分值，要求通过python编程实现。

4、综合指标权重和分值，构建暴雨内涝灾害指数评价模型。

5、利用评价模型计算城市群和各个城市的灾害指数，要求通过python编程实现。

**五、结果可视化**

1、要求制作一个或多个html网页对数据分析结果和评价结果进行可视化。