《数据科学实践》上课安排

上课时间：周五上午10:00-11:30，（3.20第一次，共16次）

上课形式：腾讯会议直播+课堂派录屏（鼓励学生直播上课，实在有困难，提前请假，课后看回放）

课程内容：一共5个大数据案例，学生分组完成，每个案例3-4周。共16名学生，每个案例，学生随机分成4组。

期末考核：考勤、平时表现、五个案例成绩的加权平均

第1周

主讲：吕晓玲

内容：介绍整体课程安排，讲解第一个案例（智能手机用户行为数据）背景、数据说明、分析要求

第2周

学生分组汇报：数据清洗、描述统计、单机实现，教师点评。

第3周

主讲：刘阳 （研究生助教）

内容：分布式处理技术，hive & spark

第4周

学生分组汇报：模型结果、分布式实现，教师点评。学生完成第一次数据分析报告。

第5周

主讲：魏太云 （业界导师）

内容：自然语言处理、文本挖掘技术在业界的应用，布置第二个案例作业（机器翻译）

第6周

学生分组汇报：数据清洗、描述统计、模型初步结果，教师点评。

第7周

学生分组汇报：模型改进结果，教师点评。学生完成第二次数据分析报告。

第8周

主讲：李舰 （业界导师）

内容：图像处理技术在业界的应用，布置第三个案例作业（图像处理）

第9周

学生分组汇报：数据清洗、描述统计、模型初步结果，教师点评。

第10周

学生分组汇报：模型改进结果，教师点评。学生完成第三次数据分析报告。

第11周

主讲：吕晓玲

内容：讲解第四个案例（美国航空数据）背景、数据说明、分析要求，布置第四个案例作业

第12周

学生分组汇报：数据清洗、描述统计、模型初步结果，教师点评。

第13周

学生分组汇报：模型改进结果，教师点评。学生完成第四次数据分析报告。

第14周

主讲：吕晓玲

内容：讲解第五个案例（纽约公共自行车数据）背景、数据说明、分析要求，布置第五个案例作业

第15周

学生分组汇报：数据清洗、描述统计、模型初步结果，教师点评。

第16周

学生分组汇报：模型改进结果，教师点评。学生完成第五次数据分析报告。