**软件工程训练营大作业要求**

软件工程训练营大作业可以从课程项目/企业项目/探索项目中任选一个，要求大作业项目具有一定的工程背景和实用价值，技术上存在一定实现难度。由小组成员分工合作共同完成该项目，每人将自己的贡献commit至相应队伍的GitHub仓库中，并在项目完成后撰写结题报告并进行答辩。根据每人在GitHub中的贡献程度，以及项目结题报告和答辩的完成情况进行评分。其中，课程推荐项目如下：

1. 基于YOLOv5和DeepSort的物体识别与跟踪

描述：小组成员自行采集视频或图片数据，结合课上所学内容，基于YOLOv5和DeepSort实现物体的识别与跟踪，并从以下任务中至少选择2个完成：

1. 基于YOLOv5和DeepSort实现视频/摄像头内的物体识别与跟踪；
2. 设计GUI界面，实时展示视频/摄像头中物体识别与跟踪的画面；
3. 能够在GUI界面中选中所有识别对象的其中一个/若干个，并跟踪被选定的对象；
4. 能够根据目标对象的移动调整摄像头/视频画面位置/放大对象所在区域，使得目标对象始终在画面中央。
5. 新能源多时空尺度出力预测

描述：小组成员根据课程提供的数据，采用包括但不限于传统机器学习与深度神经网络等方法建立数学模型，并完成以下任务：

1. 对数据进行清理，并开展探索性数据分析，包括但不限于采用图形、表格、动画等展示有用的信息，并给出初步的分析；
2. 根据分析结果抽象数学模型，并采用包括但不限于传统机器学习与深度神经网络等一种/多种方法建立初步模型，对未来7日的新能源出力进行预测，分辨率15min；
3. 分析预测结果并根据结果对模型做出改进。
4. 电力消耗、生产与天气数据分析与预测

描述：小组成员根据课程提供的数据，采用包括但不限于传统机器学习与深度神经网络等方法建立数学模型，并完成以下任务：

1. 对数据进行清理，并开展探索性数据分析，包括但不限于采用图形、表格、动画等展示有用的信息，并给出初步的分析；
2. 分析找出影响电价、电力需求、电力生产量的各种因素，分析什么样的天气会对上述量产生什么样的改变；
3. 预测24小时后的电力需求、生产和电价；
4. 评估一天内的电价与电力需求；
5. 简单评估基于当前数据下的最适合未来发展的新能源形式。
6. 小组成员根据课程提供的数据（A股股票数据），实现下列任意一个kaggle的算法，并用 qlib 或者 backtrader 对相关方法进行回测。
7. 九坤：项目目标--预测A股股票的未来一天的收益；链接：https://www.kaggle.com/competitions/ubiquant-market-prediction/discussion/338220
8. G-research：项目目标--预测加密货币的短期回报；链接：https://www.kaggle.com/competitions/g-research-crypto-forecasting/discussion/323098
9. optiver：项目目标--预测股票的波动率；链接：https://www.kaggle.com/competitions/optiver-realized-volatility-prediction/discussion/274970
10. jane-street：项目目标--制定股票交易策略；链接：https://www.kaggle.com/competitions/jane-street-market-prediction/discussion/224348
11. Two Sigma：项目目标--预测股票交易价格；链接：https://www.kaggle.com/competitions/two-sigma-financial-modeling/discussion/29793

5. 信息检索与推荐系统，小组成员根据课程提供训练数据集实现模型，验证集实现n-fold CV，最终记录在测试集表现（Metric至少需要包含AUC）提交实验流程报告。下列题目二选一：

* 1. 基于蚂蚁金融的语义匹配模型，数据集：https://tianchi.aliyun.com/dataset/dataDetail?dataId=106411
  2. 基于淘宝用户行为序列，提交一个用户兴趣模型，对用户针对商品感兴趣程度打分（购买、加购、喜欢都可以视为感兴趣标签，有pv无其他动作视为不感兴趣），11月25日-30日数据训练，12月1日-2日数据验证，12月3日数据测试。时序模型注意不要引入look ahead bias。数据集：https://tianchi.aliyun.com/dataset/dataDetail?dataId=649

6. 基于统计方法的中文词汇挖掘实践

（1） 学习了解文本数据挖掘的方法

学习路径参考

https://www.kaggle.com/code/kanncaa1/applying-text-mining

（深度学习-选做）https://www.kaggle.com/code/eliotbarr/text-mining-with-sklearn-keras-mlp-lstm-cnn

（2） 学习经典中文词汇挖掘（新词发现）方法，并使用python实现

（方法及思想）http://www.matrix67.com/blog/archives/5044

（java实现参考）https://github.com/sing1ee/dict\_build

（3）（选做）NLP领域模型发展总结，完成一个深度学习NER任务模型

实现参考（可忽略tpu部分，有条件的同学可学习使用配置nvidia-GPU）

https://www.kaggle.com/code/tanulsingh077/deep-learning-for-nlp-zero-to-transformers-bert