**《机器学习工具与平台》课程设计项目要求**

在此课程设计项目中，你将运用在此课程中学到的技能和方法选择一个感兴趣的问题来解决。回顾我们最早在《端到端的机器学习项目》实验中所经历的步骤，首先你需要定义一个你想要解决的问题，调研可能的解决方案，并解释其衡量指标。其次，你需要搜集到数据，通过数据可视化和数据挖掘分析这个问题，以便对适于解决该问题的算法和特征有一个更好的了解。接着，你需要实现你的算法和衡量指标，你需要经过数据预处理、训练模型、改善、后处理的整个过程。接下来，你需要收集这些模型的表现结果，把重要的部分可视化，来验证或证明这些结果。最后，你要在你的结果基础上得出结论，给出解决方案等，讨论一下你的实现是否真的解决了这个问题。

1. **主题**

围绕“人工智能辅助天气预报”主题，根据感兴趣的问题自行选择角度，搜集相关数据、云图、遥感卫星数据等，运用机器学习的方法对自己家乡或与自己家乡相同纬度的全球某一个城市针对天气预报相关的至少2个问题进行分析研究，应用训练所得模型结果对问题进行分析，按照模板要求完成项目报告撰写。

问题类型举例：

* 回归：以小时/天为单位划分数据，采用至少3种以上的机器学习算法（含深度神经网络），对未来某一段时间（不小于12小时）的天气情况（温度、湿度、风级、降雨量等）进行预测……
* 分类：定义天气类别（多云、晴天、雨天、雪天和雾天等），基于所查的数据集或图片集，选取至少2种以上的机器学习算法进行天气类别划分……
* 聚类：相关文献研究分析等……

数据来源举例：

* 中国气象站点数据：<https://www.cnblogs.com/icydengyw/p/12664027.html>
* KNMI Data Platform KNMI 数据平台
* Home European Climate Assessment & Dataset 欧洲气候数据中心
* https://wiki.earthdata.nasa.gov/display/GIBS NASA的地球观测数据
* ERA5-Land hourly data from 1950 to present (copernicus.eu)
* ……
* 相关论文附件或正文中所提供数据集

1. **提交报告要求**

**2.1 电子文档**

项目最终提交一份报告文件（包含封面、项目报告(正文)、代码报告三部分，合并成一个pdf）、一份代码源文件、一个 README 文档。

**封面**参考文件夹中课程设计封面。

**项目报告**按照《华东理工大学学报》论文格式要求，从问题背景、问题抽提分析到分析结论，形成一份完整的论文报告，主体部分可参考《端到端的机器学习项目》中所涉及步骤。尝试不少于4类机器学习方法，并能体现分析、改善模型的过程。运用图表体现你的过程和结果。注意，项目报告中不需要出现代码。

**代码报告**类似各堂课提交的实验报告，其中包含你全部的代码、运行结果及必要的文字说明。代码可在Jupyter或其他idle中完成，但代码需要导出为格式需为pdf格式。

**代码源文件：**上传所有代码源文件。

请将封面、项目报告、代码报告合并成一个PDF提交。

**命名格式：学号-姓名-论文名称**

**README 文档**包含你项目用到的软件以及库的描述，数据来源（链接），参考项目或论文来源（地址链接），与参考项目对比的结果和改善，说明运行你的代码所需要的（大概）时间，过程中出现的其它问题等。

README文件单独提交。

**2.2 纸质文档**

项目最终提交一份纸版报告文件（包含封面、项目报告两部分）

**2.3 报告提交时间及地点**

电子文档提交截止时间：待定

纸版文档提交截止时间：待定，地点：信息楼215

1. **其他要求**

* 独立完成报告。
* 报告中需要清楚标注数据来源、参考文献等，参考文献不少于10篇。所做项目如参考了其他文章或code项目中的方法，请清楚注明，并有适当结果对比。
* 在做课程设计的过程中，可以参考过去实验材料，尤其是端到端的机器学习项目一节。某些题目可能需要额外学习某些知识，如时间序列处理、卷积神经网络应用等，可自行寻找学习资料，也可以参考教材中相应章节内容。
* 评分标准包括选题、实现方法（所用方法、数据集是否合理且能够能够论证选题等）、运用机器学习工具分析的过程和结果、报告格式、报告易读性（结构逻辑合理、图表清晰）、结果和讨论等。