**Python机器学习和数据挖掘期未大作业**

一、要求功能：

1. 基于Web设计开发一个数据挖掘的软件，具有可操作界面，如：菜单等，可以上传数据和下载展示数据等，可以基于flask或者Django等框架；（此项是可选项，选的时候可以小组形式一起完成，不选的时候需要单独完成）
2. 可以读入一些数据集，能够对数据进行一些预处理、统计分析和可视化，能够调用这学期学过的一些机器学习方法，运用python对数据进行处理，并可以把处理结果以文档报告或者可视化图形方式表现出来。
3. 有较详细的文档说明，说明设计开发过程，如何调试运行代码以及相应效果（可以有自己录制的小视频），选用相应机器学习方法处理的实际意义等。
4. 上交时要包含可以调试运行的代码和详细文档说明（含小视频说明），另加社会调查及实际应用反馈（包括该算法软件目前在一些领域或者市场的应用具体情况和需求等）。

二、说明

（1）95分以上，达到上面的要求后，能够实现近期（近十五年）一些文章的机器学习或者数据挖掘的代码，并进行一些算法改进，附上相应的中英文文章及自己的工作意义说明，意义明显；或者软件到了一些企业或者单位实践，能够解决实际问题，得到企业或者一些单位的良好反馈；或者没有调用scikit-learn库，直接用Python实现近二十五年文章中一些经典算法，并进行一些优化改进，并应用到5个以上数据集上进行了一些运行效果比较，达到以上（2）、（3）、（4）功能（这个需要单独完成）。

（2）90-95分，达到上面的要求后，能够实现近期（近十五年）一些文章的机器学习或者数据挖掘的代码，附上相应的中英文文章及自己的工作意义说明；或者软件到了一些企业或者单位实践，能够解决实际问题，得到企业或者一些单位的肯定；或者没有调用scikit-learn库，直接用Python实现近二十五年文章中一些经典算法，并应用到5个以上数据集上进行了一些运行效果比较，达到以上（2）、（3）、（4）功能（这个需要单独完成）。

（3）75分以上，达到以上（2）、（3）、（4）要求，根据软件运行情况和文档说明情况评定（这个需要单独完成）。

（4）同时实现（1）、（2）、（3）、（4）功能的可以组1-3人小组完成。

（5）学校录入成绩结束十天前交上来，初定是6月30日，视情况可能会有变动。

三、其他评定分数具体准则：

1. 软件具有数据预处理、统计分析、可视化、机器学习及智能方法调用等功能。（50分）
2. 详细的文档说明（包含小视频说明），包括背景意义、问题及需求分析、设计及开发具体过程阐述、调试及运行测试说明（比如：需要安装什么包，如何配置和测试运行等）、效果截图展示等。（40分）
3. 社会调查及实际应用反馈等，包括企业反馈证明，或者一些单位的社会调查实践反馈及说明等（包括该算法软件目前在一些领域的应用情况和需求等）。（10分）