# 水印带英文

**《信息系统集成与开发》课程设计报告**

**姓名: 钟韩杰**

**学号: 19851260**

**班级: 计科1908**

**专业: 计算机科学与技术**

**学院: 计算机与信息技术学院**

**任课教师： 王志海**

**2022年6月16日**

# 《》课程设计报告

机器学习（Machine Learning, ML）方法的发明和应用是计算机科学中一个令人兴奋且潜在意义深远的发展。这使计*算机程序能够自动分析大量数据，并决定哪些信息最相关。这些具体化的信息然后可以用来自动做出预测或帮助人们更快、更准确地机器学习(Machine Learning, ML)方法的发明*和应用是计算机科学中一个令人兴奋且潜在意义深远的发展。这使计算机程序能够自动分析大量数据，并决定哪些信息最相关。这些具体化的信息然后可以用来自动做出预测或帮助人们更快、更准确地？？？

## 1 数据分析Weka平台网站

数据分析Weka平台网站主要分为六个部分，具体内容如下。

### 1.1 网站简介

机器学习(Machine Learning, ML)方法的发明和应用是计算机科学中一个令人兴奋且潜在意义深远的发展。这使计算机程序能够自动分析大量数据，并决定哪些信息最相关。这些具体化的信息然后可以用来自动做出预测或帮助人们更快、更准确地做出决定[[1]](#footnote-1)。

### 1.2 项目目标

本项目的只要目标有以下几个方面：

（1）使机器学习技术普遍可用；

（2）将它们应用于与新西兰工业相关的实际问题；

（3）开发新的机器学习算法并将其提供给世界；

（4）为该领域的理论框架做出贡献。

### 1.3 应用

Weka的团队已将多种标准ML技术整合到名为[Weka](https://www.cs.waikato.ac.nz/ml/weka)的软件“工作台”中，用于怀卡托知识分析环境。有了它，特定领域的专家能够使用ML从太大而无法手动分析的数据库中获取有用的知识。Weka的用户是ML研究人员和工业科学家，但它也广泛用于教学。最近，其团队还研究了[MOA](http://moa.cs.waikato.ac.nz/)，这是一种挖掘数据流的环境。

### 1.4 书籍

机器学习提供了一套令人兴奋的技术，包括分析数据和进行预测的实用工具，但也推动了人工智能的最新进展。Weka开发团队成员已经写了一本书，提供了该领域的高度可访问的介绍，但也迎合了那些想要深入研究现代概率建模和深度学习方法中可用的更多数学技术的读者。在概率模型和深度学习方面的专业知识极大地扩展了这本书的覆盖范围。为了给新材料腾出空间，现在提供了一个在线附录在Weka软件上。它是作为本书附录的Weka简要描述的扩展版本。这本书继续提供它所描述的算法的Weka实现的参考。这Weka MOOCs提供类似于第三版中的辅导练习的活动。现在还提供其他软件的信息:自从在2011年编写了第三版以来，机器学习的计算生态系统已经有了巨大的发展。第四版的目录，指出在哪里增加了新的材料，可在本页下面找到。

本书主要内容可分为：

（1）1.解释数据挖掘的机器学习算法是如何工作的。

2.帮助比较和评估不同技术的结果。

3.涵盖性能改进技术，包括输入预处理和组合不同方法的输出。

4.提供关于概率模型和深度学习的深入信息。

5.提供Weka机器学习工作台的介绍和软件中算法实现的链接。

### 1.5 课程

整合了几个免费的在线课程，使用Weka教授机器学习和数据挖掘内容。这些课程的视频可以在Youtube网站上进行观看。

1.使用Weka进行数据挖掘。

2.使用Weka进行更多数据挖掘。

3.使用Weka进行高级数据挖掘。

### 1.6 团队人员

该页面对于Weka团队成员进行了简要介绍。

## 2 系统下载、安装和运行

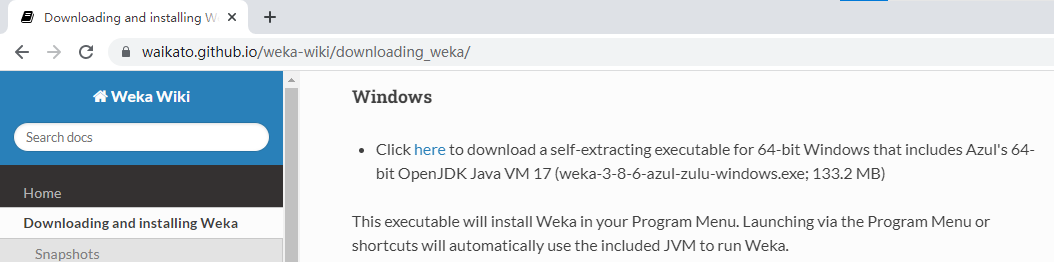
接下来将进行Weka开源软件的下载、安装与运行。

### 2.1 系统下载

进入Weka官网https://www.cs.waikato.ac.nz/ml/weka/index.html。



选择下载Windows版本的Weka 3.8.6软件。

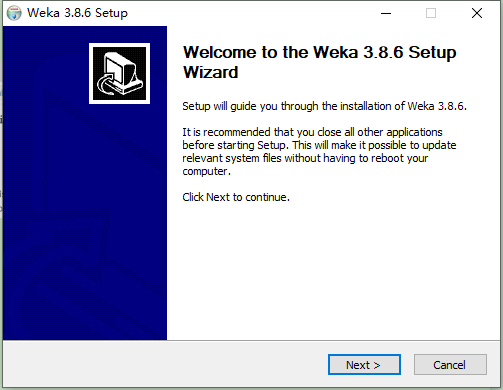


### 2.2 系统安装

下载完weka-3-8-6-azul-zulu-windows.exe文件后，即可开始安装。

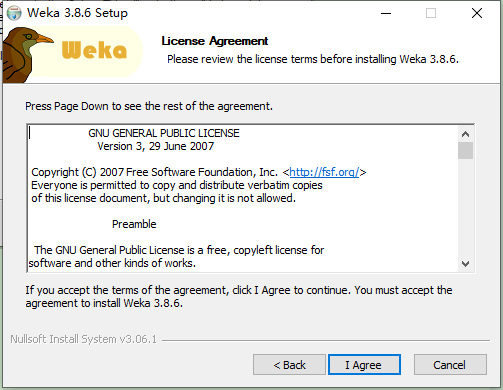
#### 2.2.1 打开安装文件

双击打开下载的weka-3-8-6-azul-zulu-windows.exe文件，点击Next进入下一步。



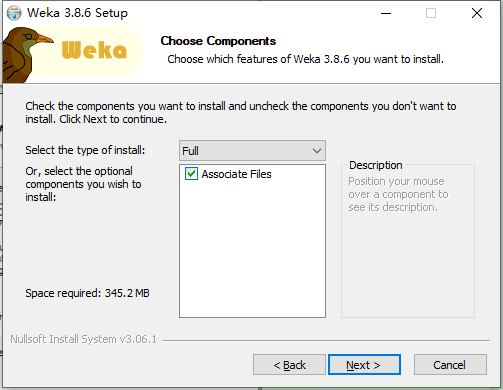
#### 2.2.2 同意使用条款

点击I Agree同意使用条款



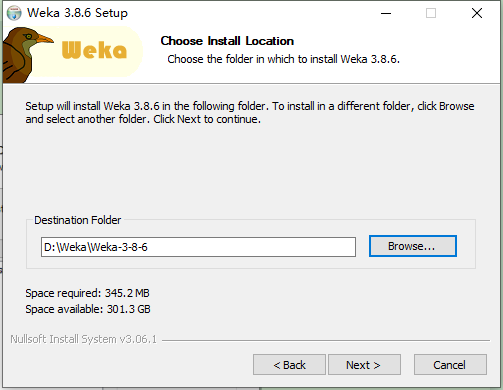
#### 2.2.3 选择安装内容

选择Full安装全部内容，并点击Next进入下一步。



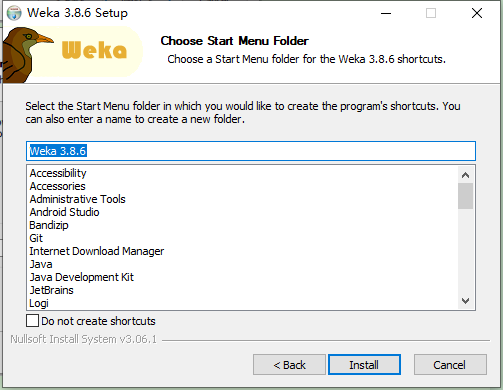
#### 2.2.4 选择安装位置

选择软件安装位置，并点击Next进入下一步。



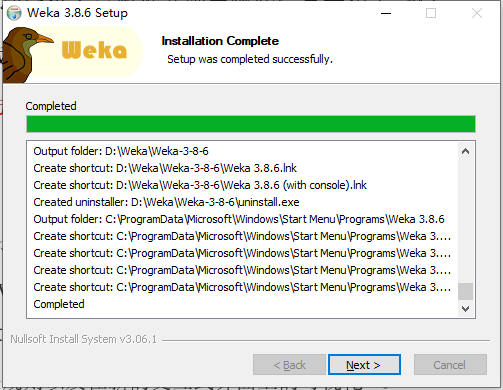
#### 2.2.5 添加到开始菜单

选择添加到开始菜单，并点击Install开始安装。



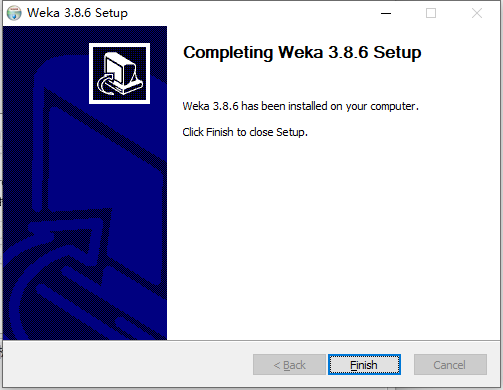
#### 2.2.6 开始安装

等待安装完毕后，即可点击Next进入下一步。



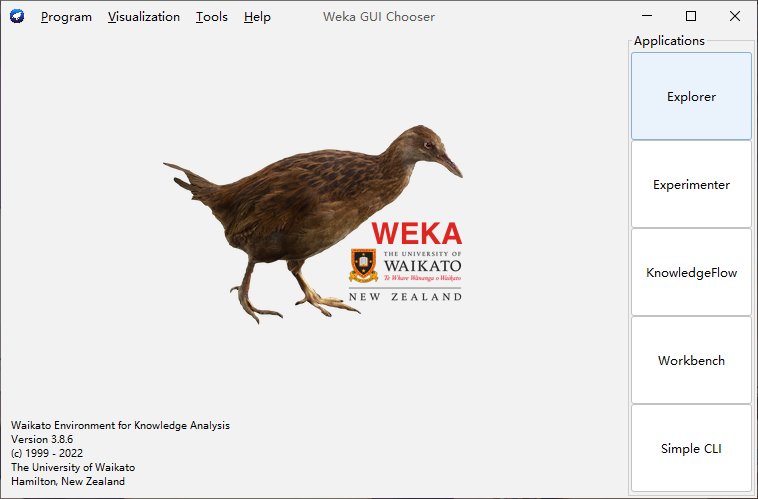
#### 2.2.7 完成安装

安装完毕，点击Finish完成安装。



### 2.3 系统运行

运行Weka 3.8.6，即可显示出如下图所示窗口。



以下是右侧各个按键功能的简要介绍：

（1）Explorer：系统提供的最容易使用的图像用户接口。通过选择菜单和填写表单，可以调用Weka的所有功能。这是用来进行数据实验、挖掘的环境，它提供了分类，聚类，关联规则，特征选择，数据可视化等等功能。

（2）Experimenter：用来进行实验，对不同学习方案进行数据测试的环境。

（3）KnowledgeFlow：可使用增量方式的算法来处理大型数据集，用户可以定制处理数据流的方式和顺序。知识流界面允许用户在屏幕上任意拖曳代表学习算法和数据源的图形构件，并以一定的方式和顺序组合在一起。按照一定顺序将代表数据源、预处理工具、学习算法、评估手段和可视化模块的各构件组合在一起，形成数据流。如果用户选取的过滤器和学习算法具有增量学习功能，就可实现大型数据集的增量分批读取和处理。

（4）Workbench：提供了一个多合一的应用程序，该应用程序包含了前面各节中描述的所有主要的WEKA GUI（Graphical User Interface，图形用户界面）。

（5）Simple CLI（command-line interface，命令行界面）：简单的命令行界面。

## 3 数据集合

Weka平台所使用的数据集文件格式为ARFF（ARtificial File Format）。接下来将借由Weka提供的weather.nominal.arff数据集进行介绍。

### 3.1 weather.nominal.arff数据集合表

表3-1 weather.nominal.arff数据集合表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | outlook | temperature | humidity | windy | play |
| 1 | Sunny | hot | high | FALSE | no |
| 2 | Sunny | hot | high | TRUE | no |
| 3 | overcast | hot | high | FALSE | yes |
| 4 | Rainy | mild | high | FALSE | yes |
| 5 | Rainy | cool | normal | FALSE | yes |
| 6 | Rainy | cool | normal | TRUE | no |
| 7 | overcast | cool | normal | TRUE | yes |
| 8 | Sunny | mild | high | FALSE | no |
| 9 | Sunny | cool | normal | FALSE | yes |
| 10 | Rainy | mild | normal | FALSE | yes |
| 11 | Sunny | mild | normal | TRUE | yes |
| 12 | overcast | mild | high | TRUE | yes |
| 13 | overcast | hot | normal | FALSE | yes |
| 14 | Rainy | mild | high | TRUE | no |

其中，各个属性的含义为：

id：取值为当前条目的序号。

outlook：取值可为sunny, overcast, rainy三种。表示天气种类。

temperature：取值可为hot, mild, cool三种。表示温度范围。

humidity：取值可为high, normal两种。表示湿度范围。

windy：取值可为TRUE, FALSE两种。表示是否有风。

play：取值可为yes, no两种。表示当天是否玩保龄球。

### 3.2 labor.arff文件数据集合

其中包含17类属性，含义如下：

1. duration：合同时间。

2. wage-increase-first-year：第一年工资增长率。

3. wage-increase-second-year：第二年工资增长率。

4. wage-increase-third-year：第三年工资增长率。

5. cost-of-living-adjustment：生活补助。

6. working-hours：工作时间。

7. pension：养老金。

8. standby-pay：未工作时工资。

9. shift-differential：换班津贴。

10. education-allowance：教育补助。

11. statutory-holidays：法定节假日数。

12. vacation：带薪假期天数。

13. longterm-disability-assistance：员工长期残疾期间的雇主补助。

14. contribution-to-dental-plan：对牙科医疗的补助。

15. bereavement-assistance：雇主对支付丧亲之痛费用的经济补助。

16. contribution-to-health-plan：对健康医疗的补助。

17. class：类别。

## 参考文献

|  |  |
| --- | --- |
| [1] | 取自Weka官网URL:https://www.cs.waikato.ac.nz/ml/weka/index.html |
| [2] |  |
| [3] |  |

1. 取自Weka官网URL:https://www.cs.waikato.ac.nz/ml/weka/index.htm [↑](#footnote-ref-1)