**

**HUNAN UNIVERSITY**

《DM&BID》

Report 1

|  |  |
| --- | --- |
| **报告名称：** | 数据挖掘与python |
| **学生姓名：** | 杨超然 |
| **学生学号：** | 202106060220 |
| **专业班级：** | 电商2102班 |
| **学 院：** | 工商管理学院 |
| **指导老师：** | 江资斌 |
| **日 期：** | 2023.1.31 |

目录

[一、知识小记 2](#_Toc439789661)

[（一） 常见的用户数据 2](#_Toc2053603100)

[（二） 大数据特征 2](#_Toc550127116)

[（三） 传统数据挖掘特征 3](#_Toc1069338277)

[（四） 数据挖掘算法 3](#_Toc77779796)

# 一、知识小记（课程刚入门，未涉及代码，故仅记录所学知识）

## 常见的用户数据

即时数据、社交数据、行为数据、用户属性数据

## 大数据特征

大数据指的是有以下四个共同特征（4V）中任意一个的数据源：**速度快（velocity）、数据量大（volume）、价值密度低（value）、类型多（variety）**

## 传统数据挖掘特征

数据准备--数据挖掘--结果表达与解释

## 数据挖掘算法

·**分类与回归**：“分类”是**对类别的预测**，即通过学习发现既有属性之间的联系，从而建立模型，对新数据的类别进行预测。“回归”是**对数值型属性的预测**

·**聚类分析**：聚类是一种在没有先验知识的条件下，根据某种相近程度的度量指标，对数据自动进行**子集划分**的技术。所形成的**子集合内部数据的结构特征相近**，不同**子集之间的数据结构特征有较大相差**

·**关联分析**：通过数据分析，找到事物之间的**相互关联规则**。包括简单关联规则和时序关联规则

·**算法结果形式**：浓缩数据、树型图、规则、数学模型

·**演绎与归纳**：演绎是从一般规律到具体现象；归纳是从具体现象到一般规律

·**有监督学习/无监督学习**：有监督**常用于分类与回归**，**目标是实现对新数据的预测**，这类算法要求数据中的**输出值是已知的**，输出就是一位“老师”，它始终监督着模型的建立和评价；无监督常用于聚类和关联分析，目标是揭示数据的内在关系和结构。由于这些关系和结构是事先未知的，所以学习过程是无“老师”参与的