

# 《商务智能与数据挖掘》课程报告

## 个人网络在线学习行为分析

姓名 1 (学号), 姓名 2 (学号), 姓名 3 (学号)

June 4, 2024

*Intentional Blank Page*

**摘要：** 这是一段用于演示在报告中显示摘要的演示内容，实际内容应根据您的研究内容总结和展示。摘要应该展示出您的研究背景、创新的研究方法以及研究结论等内容。

**关键词：** 关键词 1; 关键词 2; 关键词 3

*Intentional Blank Page*

# 目录

<b>1 引言 (问题描述)</b>	<b>1</b>
<b>2 数据预处理</b>	<b>1</b>
<b>3 实验原理</b>	<b>1</b>
<b>4 特征工程</b>	<b>1</b>
<b>5 算法选择与调优</b>	<b>1</b>
<b>6 结果分析</b>	<b>1</b>
<b>7 实验设计</b>	<b>1</b>
<b>8 实验结果</b>	<b>1</b>
<b>9 结果分析</b>	<b>2</b>
<b>10 决策支持与实际应用</b>	<b>2</b>
<b>11 结论</b>	<b>2</b>
<b>12 参考文献</b>	<b>2</b>
<b>13 商务智能与数据挖掘课程结课论文评分细则</b>	<b>2</b>
13.1 问题描述 (总分 20 分)	2
13.2 数据预处理 (总分 25 分)	2
13.3 算法选择与调优 (总分 15 分)	3
13.4 结果分析 (总分 20 分)	3
13.5 Python 编程应用 (总分 10 分)	3
13.6 决策支持与实际应用 (总分 5 分)	3
13.7 参考文献 (总分 5 分)	3
13.8 其他 (总分可视情况适当调整)	3
<b>14 文档使用示例</b>	<b>4</b>

14.1 Notebook 简介	4
14.2 文献引用示例	4
14.3 插入图片示例	4
14.4 数学公式输入示例	4
14.4.1 无编号的公式	6
14.4.2 有编号的公式	6
14.4.3 行内公式	6
14.4.4 对具有 <code>label</code> 的公式的引用	6
14.4.5 更多的公式示例	7
14.5 生成图片示例 (包含在最终 pdf 文档中隐藏不必要的 cell 内容的示例)	8
14.6 文末参考文献示例	10

## 1 引言（问题描述）

（描述研究的背景、问题、主要目标、数据来源等）

## 2 数据预处理

（应包括数据清洗、缺失值的处理、异常值检测、数据转换）

## 3 实验原理

（以动态数组与链表比较实验项目为例，简述动态数组和链表的基本原理和特点；说明为什么比较这两种数据结构是有意义的等）

## 4 特征工程

（应包括基线模型描述、效果评估，可以采用可视化的技术对效果进行详细描述）

## 5 算法选择与调优

（算法选择及调优过程，方法描述）

## 6 结果分析

（应包括模型性能评估、特征重要性、有效性等方面的讨论）

## 7 实验设计

- 说明如何设计实验来比较两种数据结构在插入、删除和随机访问操作上的性能
- 包括测试用例、测试数据和预期结果

## 8 实验结果

- 展示实验数据，可能包括图表等

- 对比动态数组和链表在各项操作上的性能

## 9 结果分析

- 分析实验结果，解释数据显示的性能差异。
- 讨论实验结果对理解这两种数据结构的特点有何帮助。

## 10 决策支持与实际应用

（应包括评估数据挖掘结果对实际业务决策支持的深度见解和实用性描述）

## 11 结论

（总结该项研究的结果和学到的知识点，改进建议的实用性和创新性）

## 12 参考文献

- 列出课程报告中引用的所有参考文献。
- 参考文献格式：GB/T7714—2015

## 13 商务智能与数据挖掘课程结课论文评分细则

### 13.1 问题描述（总分 20 分）

- 1) 相关背景（5分）：评估所选数据集的应用场景及其行业背景描述的详细性和准确性。
- 2) 问题陈述（10分）：评估问题描述的清晰度、目标明确性以及业务或组织重要性的阐释。
- 3) 数据可用性（5分）：评估数据集来源描述的清晰度和数据的适用性。

### 13.2 数据预处理（总分 25 分）

- 1) 数据清洗（5分）：评估数据清洗步骤的完整性和效果。
- 2) 缺失值处理（5分）：评估缺失值处理方法的适当性和实施效果。



- 3) 异常值检测 (5 分): 评估异常值检测方法的合理性和实施效果。
- 4) 数据转换 (5 分): 评估数据转换步骤的适用性和执行的准确性。
- 5) 特征工程 (5 分): 评估特征工程的创新性、实施效果和可视化分析的质量。

### 13.3 算法选择与调优 (总分 15 分)

- 1) 算法选择理由 (10 分): 评估算法选择的合理性、适用性描述以及理由的充分性。
- 2) 调优方法 (5 分): 评估调优方法的适当性和对模型性能改进的有效性。

### 13.4 结果分析 (总分 20 分)

- 1) 模型性能评估 (10 分): 评估使用的评估指标的适用性及其解释的深度。
- 2) 特征重要性 (5 分): 评估对特征重要性分析的深度和准确性。
- 3) 有效性讨论 (5 分): 评估解决方案有效性和局限性的讨论的全面性及其深度。

### 13.5 Python 编程应用 (总分 10 分)

- 1) 工具和库的使用 (5 分): 评估 Python 及其库的使用是否恰当、有效。
- 2) 代码片段 (5 分): 评估代码的清晰度、注释的充分性和代码的功能性。

### 13.6 决策支持与实际应用 (总分 5 分)

- 1) 业务决策支持 (3 分): 评估数据挖掘结果对实际业务决策支持的见解深度和实用性。
- 2) 改进建议 (2 分): 评估改进建议的实用性和创新性。

### 13.7 参考文献 (总分 5 分)

- 1) 参考文献 (5 分): 评估参考文献的广度、相关性和正确引用。

### 13.8 其他 (总分可视情况适当调整)

- 1) 论文结构 (5 分): 评估论文的整体结构、逻辑性和清晰度。
- 2) 论文表达 (5 分): 评估论文的语言表达、图表使用的恰当性和可读性。
- 3) 文件组织 (5 分): 评估提交的文件组织结构和命名的规范性。

## 14 文档使用示例

### 14.1 Notebook 简介

Jupyter notebook 的 Cell 可以是 Markdown、Code 以及 Raw 状态。不同的状态其功能显然不同。比如在 Markdown 状态下，该 Cell 支持 [CommonMark](#) 标准的 Markdown 语法。在 Code 状态下，可以输入相应的代码，比如 Python、R、Matlab 等。如果要运行相应的代码，需要确认是否安装了相应的内核，特别是除 Python 以外的其他语言。在本门课程中，Markdown 主要用于对课程报告的背景、数据预处理过程等等的描述。其基本的使用过程，请同学们参考 [CommonMark](#) 标准，或者 [Basic writing and formatting syntax](#)，以及 [GitHub Flavored Markdown Spec](#)。关于如何使用 Jupyter notebook 来做数据分析，请参考 [Jupyter notebook](#) 官网的介绍

### 14.2 文献引用示例

关于文献引用的问题，该课程推荐使用 [Zotero](#) 文献管理器，并结合其 [Better BibTeX](#) 插件来导出格式为 `.bib` 的文件。然后将其放到与您的 `notebook.ipynb` 文件所在的文件夹。这样我们就可以使用 latex 文档中引用文献的基本范式在正文中插入相应的文献。比如，在一句话的末尾使用 `\citep{tapioTheoryDecouplingDegrees2005}`，然后使用 `jupyter nbconvert notebook.ipynb --to pdf --template=coursereport` 命令将整个文档转换为 pdf 格式时，会在 pdf 中显示为：“这是一段示例 ([Tapio, 2005](#))”。这样的标准引用格式，并且在文末生成相应的参考文献，且符合 GBT7714-2015 的标准的文献格式。如果您想表达“XXX 的研究阐明了一个 YYY 的科学事实”，你可以使用“`\citet{shaikCropInsuranceAdjusted2013}` 的研究阐明了一个 YYY 的科学事实”这样的方式来引用相关文献。在 pdf 中会显示为：“[Shaik \(2013\)](#) 的研究阐明了一个 YYY 的科学事实”。更为具体的介绍可以查看 `natbib` 包的说明。

### 14.3 插入图片示例

插入图片遵循 Markdown 插入图片的基本语法 (``)，具体的方式建议查看 Markdown 的基本使用。

### 14.4 数学公式输入示例

该模板使用了 `amsmath` 包来渲染数学公式，具体的公式编辑格式可以参考 [overleaf](#) 网站上的 Mathematics 模块的介绍。以下是可能用到的公式编辑示例：

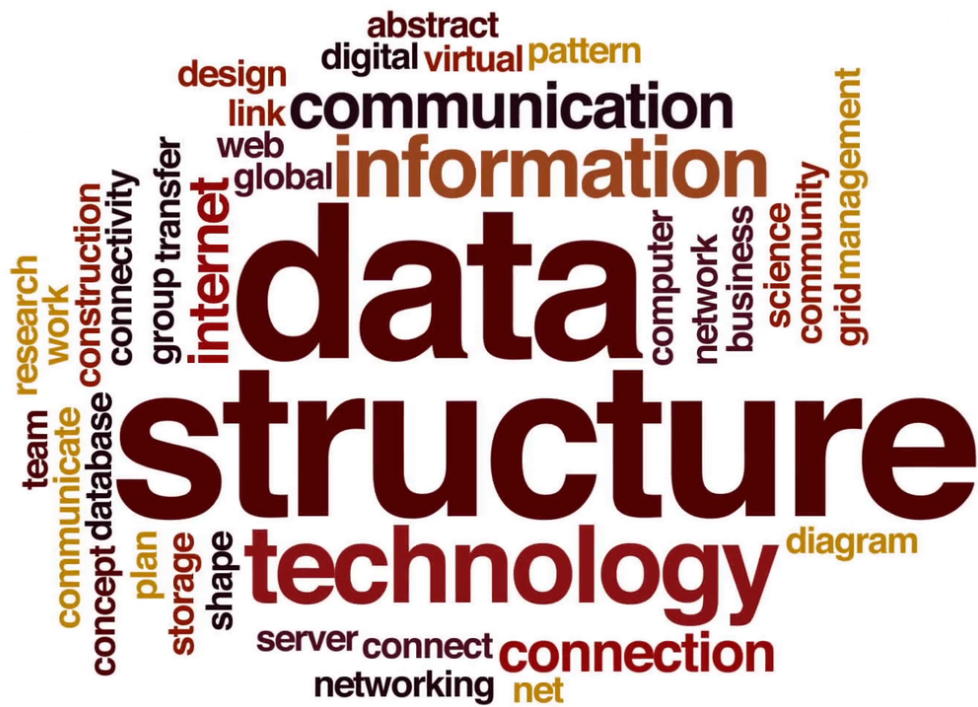


图 1: 这是插入图片的标题示例



Warning: No caption specified. Please set a caption in the notebook metadata.

### 14.4.1 无编号的公式

\$\$

$$y = \sum_{i=1}^n i$$

\$\$

以上代码将生成如下的形式：

$$y = \sum_{i=1}^n i$$

### 14.4.2 有编号的公式

```
\begin{equation}\label{eq:sum}
```

$$y = \sum_{i=1}^n i$$

```
\end{equation}
```

以上代码块将显示为：

$$y = \sum_{i=1}^n i \tag{1}$$

### 14.4.3 行内公式

这是一段描述公式内容的语句，在这插入一个行内公式  $y + x = z$

### 14.4.4 对具有 label 的公式的引用

这是一段描述公式内容的语句，如公式 (1) 所示。

关于在 Notebook 中输入公式，仍然需要参考 Markdown 的基本语法，如果想生成可以引用的公式，需要给公式添加上 label。由于 Notebook 使用了 Katex 来渲染公式，这个时候在 Notebook 中会显示解析错误，不过没有关系，在使用命令转换时，会在 pdf 中正确显示（只要您的公式符合 latex 的基本语法）。

## 14.4.5 更多的公式示例

$$\begin{aligned}
 A &= \frac{\pi r^2}{2} \\
 &= \frac{1}{2}\pi r^2
 \end{aligned}
 \tag{2}$$

- Displaying long equations

$$p(x) = 3x^6 + 14x^5y + 590x^4y^2 + 19x^3y^3$$

$$-12x^2y^4 - 12xy^5 + 2y^6 - a^3b^3$$

$$2x - 5y = 8$$

$$3x + 9y = -12$$

- Aligning several equations

$$x = y$$

$$w = z$$

$$a = b + c$$

$$2x = -y$$

$$3w = \frac{1}{2}z$$

$$a = b$$

$$-4 + 5x = 2 + y$$

$$w + 2 = -1 + w$$

$$ab = cb$$

- Grouping and centering equations

$$2x - 5y = 8$$

$$3x^2 + 9y = 3a + c$$

以上公式的源代码请查看对应的 notebook 文件。

### 14.5 生成图片示例 (包含在最终 pdf 文档中隐藏不必要的 cell 内容的示例)

表 1: 这是表格的标题

Pclass	Name	Sex	Age	SibSp	Parch
0	3 Braund, Mr. Owen Harris	male	22.000000	1	0
1	1 Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.000000	1	0
2	3 Heikkinen, Miss. Laina	female	26.000000	0	0
3	1 Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.000000	1	0
4	3 Allen, Mr. William Henry	male	35.000000	0	0
5	3 Moran, Mr. James	male	NaN	0	0
6	1 McCarthy, Mr. Timothy J	male	54.000000	0	0
7	3 Palsson, Master. Gosta Leonard	male	2.000000	3	1
8	3 Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.000000	0	2
9	2 Nasser, Mrs. Nicholas (Adele Achem)	female	14.000000	1	0

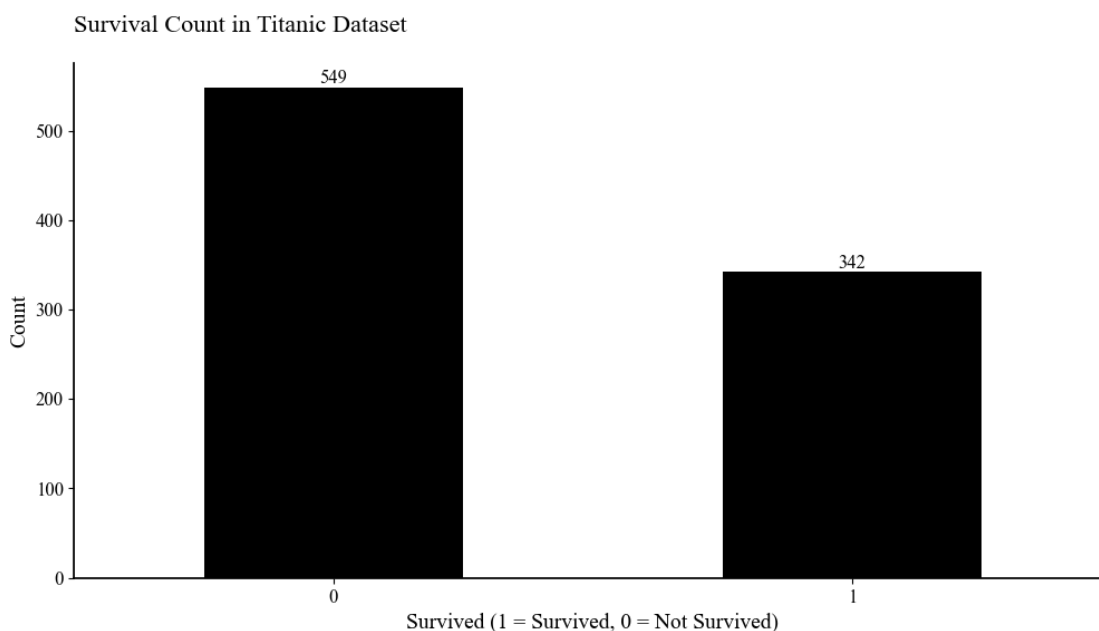


图 2: 这是生成图片的标题示例

```
# 假设我们想展示 survival_counts 和 survival_rates 的数据,
# 考虑简便, 我们可以直接使用 print 打印出相应的内容
# 但是, 此时通常会显示建议将其转换成 DataFrame 格式显示, 效果会更好
# 因此, 请一定采纳这个建议。
```

```
print(survival_counts)
```

```
print(survival_rates)
```

Survived

0 549

1 342

Name: count, dtype: int64

Survived

0 61.62

1 38.38

Name: proportion, dtype: float64

**Warning: Consider converting this output to a DataFrame format for better data manipulation.**

表 2: 这是经过转换后的表格的标题

	Counts	Rates
Survived		
0	549	61.620000
1	342	38.380000

# 以下会输出错误提示, 你应该解决这个问题, 不要在最终 *pdf* 文档中出现。如果没解决, 你会在生成的 *pdf* 文档中看见警告

```
print(survival_countss)
```

```
print(survival_ratess)
```

**Warning: There is an error in the output, please check your code carefully.**

代码 1: 这是代码块的标题

```
# 代码片段展示示例
```

```
class DynamicArray:
```

```
    def __init__(self):
```

```
self.array = [] # 初始化一个空列表，作为动态数组的底层数据结构
def add(self, item):
    self.array.append(item) # 添加元素到数组的末尾
def delete(self, index):
    if index >= len(self.array) or index < 0:
        print("Index out of bounds") # 检查索引有效性
        return
    self.array.pop(index) # 删除指定索引处的元素
def get(self, index):
    if index >= len(self.array) or index < 0:
        print("Index out of bounds") # 检查索引有效性
        return None
    return self.array[index] # 访问指定索引处的元素
def __repr__(self):
    return str(self.array) # 返回数组的字符串表示，方便打印
```

引用以上代码片段示例：如代码 (1) 所示。

### 警告：请仔细检查生成的 pdf 文档，至少确保没有紫红色的 Warning

最后需要说明的是，该文档转换为 pdf 后是支持双面打印的格式，提交课程报告时，如老师强调提交纸质版，请直接双面打印。

## 14.6 文末参考文献示例

### 参考文献

TAPIO P. Towards a theory of decoupling: Degrees of decoupling in the EU and the case of road traffic in Finland between 1970 and 2001[J]. Transport Policy, 2005, 12(2): 137-151.

SHAIK S. Crop Insurance Adjusted Panel Data Envelopment Analysis Efficiency Measures[J]. American Journal of Agricultural Economics, 2013, 95(5): 1155-1177.