**课程论文大纲及要求**

**课程名称【机器学习基础】**

**课程英文名【 Statistics case practice】**

**1. 题目与摘要**

自选一个与机器学习相关的实际应用问题作为研究主题。提供一个简洁的题目，准确反映研究内容，并撰写一个包含关键信息的摘要，字数一般不超过300字。

**2. 引言**

- 介绍所选问题的背景、重要性及其在现实世界中的应用。

- 阐述研究问题、目标和预期成果。

**3. 数据收集与预处理**

- 描述数据来源、数据集的特点以及数据预处理的步骤。

**- 训练数据集的样本量需要在5,000以上。**

参考数据来源：

* + <https://archive.ics.uci.edu/datasets>
  + <https://www.kaggle.com/datasets>
  + <https://catalog.data.gov/dataset>

**4. 机器学习模型选择与描述**

- 解释为什么选择特定的机器学习模型，详细描述模型的理论基础。

- 如果未对特定模型原本的设计参数（例如卷积核大小、通道数等），解释为什么原本的设计参数能够满足当前所研究的问题以及训练数据。

- 如果调整了特定模型原本的设计参数，解释这么调整的原因。

- 详细说明实验设计，包括数据分割、损失函数、模型训练算法和评估方法等细节。

**5. 实验结果与分析**

- 展示实验结果，使用图表和文字进行详细分析。

- 对比不同模型或参数调整的结果。可以与特定模型原本的设计参数的运行结果进行比较，或与线性回归、逻辑回归、softmax回归等较为简单的模型结果进行比较。

**6. 结论**

- 分析实验结果，讨论模型的性能、优点和局限性。

- 探讨可能的改进方向和模型在实际应用中的潜在影响。

- 总结研究的主要发现，强调研究的贡献和实际应用价值。

**7. 参考文献**

- 列出所有引用的文献，确保遵循适当的引用格式。

**8. 格式与提交要求**

- 组队要求：可以个人单独完成期末作业，也接受以2人小组提交的作业。

- 字数限制: 正文部分建议字数在3000-5000字之间。

- 文档格式：要求提交Word或PDF格式的文档，使用10.5号的宋体，英文使用Times New Roman字体，双倍行距。

- 模型实现：必须使用至少一种机器学习库或工具（如scikit-learn, TensorFlow, PyTorch等）来实现所选模型。

- 代码展示：代码应作为论文附录提交。代码应确保提供足够的信息，使得他人能够复现实验结果。

- 截止日期：2024年6月30日中午12:00前提交。

**9. 评价标准**

* 内容质量（40%）。包括：a)问题阐述的清晰度；b)对现有研究的理解和分析的深度；c)模型选择的理由及其适用性；d)实验设计的严谨性；e)结果的准确性以及分析的深度。
* 方法论（20%）。方法论的合理性：所采用的方法和步骤是否科学、系统。模型构建：模型实现的技术细节和准确性。结果复现：提供的资料是否足够让他人复现实验。
* 结构与组织（15%）结构清晰度：论文是否有逻辑清晰的结构，各部分是否有序排列。流畅性：论述是否连贯，段落之间是否有良好的衔接。格式规范：是否遵循了给定的格式要求，如字体大小、行距、页边距等。
* 引用规范（10%）：所有引用的文献是否按照规定的引用格式正确列出。
* 写作质量（10%）：语言是否清晰、准确、无语法错误。专业术语的使用是否得当，是否对关键术语进行了定义。
* 创新性与深度（5%~25%）：论文是否有创新点。如果创新性较强，将适当额外加分。

**10. 学术诚信**

**- 不允许出现不同小组采用相同数据集以及相同模型的情况。如果出现该情况，所有出现雷同的小组期末成绩直接为0。**

**- 工作必须是原创的，如若发现严重的抄袭现象，期末成绩直接为0。**