

肖智元

意向岗位: 数据分析师

电话: 18527850256

邮箱: 1786032122@qq.com

性别: 男

年龄: 21



教育经历

2020.09 - 2024.06

西华师范大学

人工智能 | 本科

科学计算与数据分析、数据库原理及应用、线性代数、概率论与数理统计、数据结构、离散数学、操作系统。



相关技能

- 数据库: 熟悉MySQL
- 数据处理和分析: 熟悉Numpy、Pandas、Matplotlib
- 网络数据爬取: 熟悉Requests、Selenium、BeautifulSoup4、Xpath、正则表达式
- 数据清洗和分析工具: 熟悉Excel (基本函数和Power Query)
- 可视化工具: 熟悉PowerBI
- 机器学习算法: 了解Logistic Regression、支持向量机、K最近邻分类、高斯朴素贝叶斯、分类决策树、随机森林分类



项目经历

2022.06 - 2022.06

课程数学建模结课项目

背景: 数学建模校赛《三胎政策对人口状况、经济及其影响的研究》

相关分析工具以及模型: matlab, python, numpy, sklearn, Leslie模型

任务描述:

- 2035年时中国的人口结构情况。
- 分析三胎政策对国家经济增长情况的影响。

项目过程:

- 分析问题选择模型:** 在开始项目之前, 我进行了广泛的研究并阅读了相关文献。最终, 我选择了Leslie模型作为研究人口结构和经济增长的分析工具, 因为它在这个领域被广泛使用且具有较高的准确性和可靠性。
- 收集数据与数据清洗:** 研究人口结构以及经济情况, 选择《中国人口统计年鉴》上收集到的1980年到2020年部分数据, 总人口、出生人口、性别比例、自然增长率、GDP、结婚率共六列数据作为特征值进行收集。在数据收集后, 我进行了数据清洗工作, 对缺失值进行了补充, 并确保数据的准确性和完整性。
- 数据分析:** 基于Leslie模型, 我建立了一个离散模型, 将人口划分为92个年龄段(0岁到91岁)。针对每个年龄段, 我分析了存活率、生育率(结婚女性)、出生率(男)、出生率(女)和劳动力等指标。通过

前一年的数据预测，我得出了每年的人口结构和经济增长情况。

项目结果与分析：通过Leslie模型，建立了关于人口的离散模型，在社会稳定的情况下对人口结构与经济增长情况进行了分析与预测，在预测中使用了第六次人口普查的数据，使模型计算结果更加准确。预测到2035年，中国的人口结构将呈现出逐渐老龄化的趋势，其中60岁及以上的老年人口数量将大幅增加，而劳动力人口比例可能会下降。然而，三胎政策有望缓解劳动力短缺问题，并通过增加新生儿人口数量来提供更多潜在的劳动力资源。这将有助于稳定和促进国家的经济增长。

2022.09 - 2022.09

对古代玻璃制品成分的分析与鉴别

数学建模与编程

背景：

在2022年的数学建模省赛B题中，我作为团队的成员，对问题三四进行建模分析预测

相关分析工具与模型：python、numpy、pandas、sklearn、SPSS、决策树、Pearson 相关性

任务：

- 问题三：未知类别玻璃文物的化学成分进行分析，鉴别其所属类型，并对分类结果的敏感性进行分析。
- 问题四：针对不同类别的玻璃文物样品，分析其化学成分之间的关联关系，并比较不同类别之间的化学成分关联关系的差异性。

数据处理：对数据进行处理得到纹饰、类型、颜色、表面风化、以及二氧化硅(SiO₂)、氧化钠(Na₂O)等十四种化学成分

问题三行动：

- **分析问题选择模型：**根据文献阅读和调研以及相关算法的分析，选择有监督的决策树。
- **数据分析：**通过有监督的决策树算法，先对已知类别数据进行训练搭建模型，最后将未知类别数据代入分类。由于当样本数量少并且样本特征多时，决策树容易出现过拟合情况，因此在训练模型之前，先进行维度规约即主成分分析来降低特征维度。由于本文的分类回归树划分不是很多时，不会对噪声数据产生过拟合，因此不需要通过剪枝来解决。最后将未知类别数据代入模型进行预测得到分类结果。决策树算法作为一个有监督的，并且更能挖掘出准确性高的分类规则的算法。

结果：解决对分类结果的敏感性分析，需要从分类模型的评价指标：准确率（Accuracy）、精确率（Precision）、召回率（Recall）进行具体分析。

模型效果评估：模型与本文的数据非常契合，预测分类的结果很好。由于之前在训练模型中为了防止出现结果过拟合现象，就进行了数据维度规约，并且在训练模型过程中进行了剪枝操作。因此，该模型很稳定，分类结果准确。

问题四行动：

- **分析问题选择模型：**数据为连续变量，因此选择 Pearson相关性分析多变量间的关系，使用偏相关分析更进一步分析两两变量之间的相关关系并判断化学成分关联关系的差异性的差异性。
- **数据分析：**对数据进行决策树分析。可以得到，氧化铝作为区别高钾玻璃和铅钡玻璃中最关键化学成分，再应用偏相关分析可在排除了其他变量的影响情况下，再计算两变量的相关关系

结果：得到了不同类别的玻璃文物样品，分析其化学成分之间的关联关系，并比较不同类别之间的化学成分关联关系的差异性

模型效果评估：决策树算法更适应与这样的小数据集，可以处理不相关特征数据。并且效率高，便于解释，决策树只需要一次构建，反复使用，每一次预测的最大计算次数不超过决策树的深度。

2021.10 - 2022.07

ELE英语学习网站

用户行为分析

背景：

对校内英语学习网站项目产生的用户信息进行分类分析，用以监督、帮助大家英语学习

相关工具：SQL、python

任务：对不同的用户制作通用的学习看板，可视化用户的学习进度

行动：

- 数据提取：使用SQL和Python对不同用户进行分组，并收集他们每天的学习数据，以了解他们的学习进度。
- 数据建模：根据艾宾浩斯遗忘曲线的原理，利用用户的学习数据建立了一套学习方案。根据该曲线，我们设计了间隔为1天、2天、5天、8天和14天的学习复习计划，帮助用户优化记忆效果。
- 数据可视化：针对不同用户，我们制作了学习进度和学习看板。通过直观的图表展示，用户可以清楚地了解自己的学习进度，并且可以及时调整学习计划。

成果：

- 提高学习效率：根据用户的反馈和学习数据分析，我们发现使用学习看板的用户相比未使用者，学习效率提高了30%。
- 增强记忆效果：通过定期复习和艾宾浩斯遗忘曲线的学习方案，用户的记忆效果明显提升。例如，在使用我们的学习方案后，用户的记忆保持率在30天内高于传统学习方法的平均值。
- 提供个性化建议：通过学习看板的数据分析，我们能够给出针对每个用户的个性化学习建议，使他们能够更好地利用时间和资源，提高学习效果。



博客

CDSN：<https://blog.csdn.net/sjyrpb>

项目主页：<https://sjyrpb.github.io>



在校经历

2020.06 - 2024.09

比赛

数学建模四川省省一等奖

蓝桥杯B组省二等奖

计算机设计大赛校二等奖



自我评价

我有数量统计技能、编程技术能力以及问题解决能力。注重细致入微和准确性，在数据处理过程中力求精确性。具备建模和预测能力，能够使用统计方法和机器学习技术构建模型并进行预测。还具备良好的沟通能力，能够将复杂的数据分析结果以简明扼要的方式传达给非技术人员。



实习经历

开始时间 - 结束时间

公司名称

职位名称

详细描述你的职责范围、工作内容和工作成果。最新的工作经验放在最前，描述尽量简洁，尽量写与职位相匹配的内容，将有助于HR第一时间发现你的亮点。如果是不知名企业，可以在工作内容的第一句话加上简短的公司或主要产品介绍，尽可能用数字说明成绩，突出分析能力、团队协作能力、解决问题的能力等工作上所需的专业素质。