

刘思忆

chaoxisuoding@outlook.com

微信: fernweh62

个人介绍

电话: 13882912003



我的主修课程与大数据相关，对数据分析和处理有着扎实的基础。同时，我是一名热爱游戏的玩家，喜欢在PC端steam和SwitchNS上玩放松、推理类的游戏。我对游戏的理解不仅停留在娱乐层面，更注重分析游戏机制和用户体验。

欢迎访问我的个人介绍网站: <https://silvia-iota.vercel.app/en>

统招全日制 统计学本科 “双一流”建设高校 读书社 组织部部长 2021.10-2024.06

东北农业大学 (211) 06/2025毕业 职责: 规划活动、编写并审核策划、记录项目进展、协调活动执行

- 精通 Python、R语言、C语言和 JavaScript等编程语言
- 熟练运用MySQL进行数据库的设计和管理
- 熟悉Github、Git的版本控制和Linux常用命令
- 数字化工具: Tableau、Microsoft Office (Excel, Word, PowerPoint)、Power Query、数据透视表、SPSS
- 机器学习: 随机森林/逻辑回归/时序/KNN/K-means/聚类模型处理分群/分类模型处理预测/假设检验ab

实习经历 海南亿启房地产有限公司 数据科学实习生 2024.05-2024.09

- 运用机器学习技术进行建模、模型工程解决数据科学中包括增效度量、因果推断等问题。
- 揭示数据与业务之间的关联，利用数据进行业务建模和定量推理，为公司**增加了28.5%**的咨询量。
- 数据的获取、清洗、建模和探索，涉及技术工具主要为SQL、Excel和Python，并制作可视化图表和报告。

项目: 竞赛论文的辅助自动评阅 “泰迪杯”数据挖掘赛挑战赛 **全国三等奖** 2024.04-2024.05

描述: 我作为队长兼编程手，负责用Python和R语言开发论文质量评价系统，该大语言模型通过综合使用数据挖掘、自然语言处理、文本分析、机器学习，从五个维度为论文评分，最后获得全国三等奖。

- 特征工程: 遍历pdf, 用正则找 (正文和摘要)、预处理、定义评价函数、设置异常时的条件函数和评分梯度。
- 主题建模: 训练模型，利用LDA (潜在狄利克雷分布) 方法从文本中提取主题，判断是否与比赛题目相关。
- 竞赛成果: 该模型共计为**810篇**论文批量式打分，并且实际拟合效果的匹配度**高达91.3%**，效率**提高近300%**。

项目: Femicide数据可视化 2023.10-2023.12

- 数据清洗和预处理: 运用Numpy、Pandas对原始数据进行全面的清洗和预处理，确保数据的准确性和一致性，包括处理缺失值、异常值以及数据格式转换。
- 数据可视化: 使用例如Matplotlib、Seaborn、Polly等库，直观地展示Femicide数据中的关键趋势和特点。通过这些可视化，用户可以深入了解数据背后的故事。

证书

CET-4	2021.12
驾驶证C1	2022.02
泰迪杯数据挖掘赛 (大语言模型)	全国三等奖 2024.06



Siyi Liu

Phone number: (+86) 13882912003 (Mobile) | **Email address:**

chaoxisuoding@outlook.com | **Website:** <https://silvia-iota.vercel.app/en> |

WeChat: fernweh62

● EDUCATION AND TRAINING

09/2021 – 07/2025 Harbin, China

BACHELOR OF SCIENCE IN STATISTICS Northeast Agricultural University (211)

● WORK EXPERIENCE

05/2024 – 09/2024

DATA SCIENTIST HAINAN YIQI REAL ESTATE COMPANY

- Utilized ML techniques for modeling and model engineering to solve data science problems.
- Uncovered relationships between data and business, employing data for business modeling and quantitative reasoning, which led to a 28.5% increase in consultations.
- Engaged in data acquisition, cleaning, modeling, and exploration, primarily using SQL, Excel.

● DIGITAL SKILLS

Python | MySQL | C Language | HTML/CSS | Flourish | Rstudio | SPSS | Data mining | TABLEAU | GIT version control, Linux Command | Machine Learning, NLP | Good English efficiency skills | Power BI/ Power Pivot | Microsoft Office (Outlook, Excel, Word, PowerPoint)

● PROJECTS

10/2023 – 12/2023

Femicide Data Visualization

- **Data Cleaning and Preprocessing:** Applied Numpy and Pandas for thorough data cleaning and preprocessing, ensuring accuracy and consistency, including handling missing values, outliers, and data format conversions.
- **Data Visualization:** Utilized libraries like Matplotlib, Seaborn, and Plotly to visually display key trends and characteristics in the Femicide data.

● HONOURS AND AWARDS

Data Mining Paper Quality Evaluation System National Third Prize

As the team leader and lead programmer, I developed this language model using Python and R, utilizing data mining, NLP, text analysis, and ML, ultimately earning a national third prize.

- **Feature Engineering:** Scanned PDFs using regular expressions to identify text and abstracts, performed preprocessing, defined evaluation functions, and set conditional functions and grading scales for exceptions.
- **Topic Modeling:** Trained models using LDA (Latent Dirichlet Allocation) to extract themes from text and assess topic relevance to the competition's theme.
- **Competition Results:** The model scored 810 papers in batches, achieving a match rate of 91.3% with actual results and a nearly 300% increase in efficiency.

CET-4