

蔡依珊

手机：13341182001 | 邮箱：cys021031@163.com | 微信：a3251907



教育背景

加州大学圣地亚哥分校，数据科学，数学-概率与统计，本科 2021.09 - 2025.06

- GPA: 3.84 / 4.0; 荣誉：教务长荣誉奖
- 相关课程：数据结构与算法，机器学习，数据挖掘，数据库管理，数据可视化，概率论，数理统计，随机过程，金融数学，应用线性代数，多元微积分

技能

编程技能：Python, SQL, Java, Spark, R, MATLAB; JavaScript, HTML, CSS

工具：Git, AWS, VS Code; Tableau, Power BI, Excel, Stata, ArcGIS

语言：英语（托福 106）

实习经历

腾讯科技（北京）有限公司 2024.06 - 2024.09

CSIG 事业群 - 企点产品部 - 商业分析实习生 中国，北京

- 参与企点营销云中的客户数据平台（CDP）、营销自动化平台（MA）、和会员管理平台（LM）的优化
- 开发 ETL 流程实现数据的提取、验证和转换，通过 SQL 预处理原始事件数据并构建标签体系，支撑日活 15 万+ 用户的行为数据处理，为营销决策提供分析支持
- 分析转化漏斗，追踪从应用启动到支付的全流程，识别关键流失点，并提出优化策略，使用户激活率提升 10%
- 通过 Tableau 构建自动化数据看板，监控不同生命周期人群的关键指标，识别高价值用户群（前 20% 用户贡献 40% 营收），并制定个性化营销策略
- 展开会员体系竞品分析，设计并优化会员管理机制，提升用户粘性与复购率

第四范式科技有限公司 2023.06 - 2023.09

科学技术部 - 大模型数据实习生 中国，北京

- 整理总结小红书笔记、知乎文章、今日头条文章等平台的文本数据特征，利用 Python 进行大规模非结构化数据清洗，处理总量达 923G
- 对清洗后的语料库数据进行准确性、完整性和规范性的抽样评估，完成 45 份 + 针对数万条数据的评估报告，优化后平均准确率提高约 22%
- 为不同任务类型构建模板，包括中英翻译、文本摘要、情感分析、阅读理解、信息抽取等，创建 2000+ 个 prompts，并编写监督式微调（SFT）模型训练的答案以提升模型性能

项目经历

文本情感分析与情感转换，组员 2024.09 - 2024.12

- 通过 GPT 生成并人工优化，构建包含 6 种情感、10 个强度等级的 3,000+ 句情感数据集
- 基于 BERT 预训练模型构建文本情感分类系统，引入自注意力机制捕获文本的上下文依赖特征，通过交叉熵损失函数和 AdamW 优化器进行模型训练，实现 98% 的验证准确率
- 构建 BERT 回归模型用于情感强度预测，并结合掩码与缩放技术，将均方误差降低至 0.1
- 设计基于表征工程与 T5 Transformer 的情感转换方法，可根据目标情感调整句子情绪，通过人工评估验证效果

远程工作与心理健康的因果发现分析，组长 2024.09 - 2024.12

- 主导因果发现算法（PC、FCI、GES）性能分析，研究远程办公对心理健康的影响
- 设计包含线性与非线性关系的模拟数据，涵盖社交孤立、工作时长、压力等关键变量
- 采用 Fisher Z、KCI、ANM 等方法，实现 87.5% 的因果推断准确率，验证远程办公与心理健康的分析框架

银行开户欺诈检测，联合负责人 2024.01 - 2024.03

- 分析银行账户欺诈数据集，进行探索性数据分析（EDA）、统计检验、特征选择，并使用 Matplotlib 进行可视化
- 应用及优化决策树、XGBoost 和 CatBoost 算法的预测模型，以检测欺诈性银行账户申请，达到 80% 的准确率
- 进行特征工程和超参数调优，使用 GridSearch 进行 5 倍交叉验证，从而识别出对模型性能提升重要的关键特征

全球互联网使用趋势可视化，组员 2024.01 - 2024.03

- 使用 Svelte、HTML 和 JavaScript 开发网站结构和样式
- 设计全球互联网使用趋势的交互式地图，伴随集成滑动条、搜索功能、排名列表和动态折线图交互组件，提升用户体验和数据分析能力

Yishan Cai

yic075@ucsd.edu | (858) 247-9139 | [LinkedIn](#) | La Jolla, CA

EDUCATION

University of California, San Diego

La Jolla, CA

B.S. in Data Science & B.S. in Probability and Statistics

Sep. 2021 – Jun. 2025 (expected)

- **GPA:** 3.84/4.0, Provost Honors
 - **Relevant Coursework:** Data Structure and Algorithms, Machine Learning, Data Mining, Database Management, Data Visualization, Scalable Analytics Systems, Math Statistics, Stochastic Process, Multivariable Calculus, Linear Algebra
-

SKILLS

- **Languages:** Python, SQL, Java, Spark, R, MATLAB, JavaScript, HTML, CSS
 - **Libraries & Tools:** NumPy, Pandas, Scikit-learn, PyTorch, Tensorflow, Matplotlib; AWS, Tableau, Power BI, Stata, ArcGIS
-

WORK EXPERIENCE

Tencent Technology (Beijing) Co., Ltd

Beijing, China

Business Analyst Intern, Qidian Product Department

Jun. 2024 – Sep. 2024

- Participated in the optimization of the Customer Data Platform (CDP), Marketing Automation Platform (MA), Loyalty Management Platform (LM) for Qidian Marketing Cloud, a B2B SaaS Intelligent Marketing Platform
- Developed an ETL pipeline to ingest, validate, and transform data for 150K+ DAU, using SQL query for preprocessing raw event data and computing user behavioral metrics
- Conducted conversion funnel analysis from app launch to payment, identifying key drop-off points and proposing optimization strategies that improved user activation rate by 10%
- Built automated dashboards in Tableau for user segmentation analysis, identifying high-value user segments (top 20% users contributing 40% revenue) and developing personalized marketing strategies
- Initiated competitive analysis on membership strategies and designed an enhanced membership system

Fourth Paradigm Technology Co., Ltd.

Beijing, China

LLM Data Intern, Science and Technology Department

Jun. 2023 – Sep. 2023

- Compiled and summarized text data characteristics from platforms such as Xiaohongshu, Zhihu, and Toutiao, and performed large-scale unstructured data cleaning using Python, handling a total volume of 923 GB
 - Executed sampling evaluations of the cleaned corpus for accuracy, completeness, and consistency, produced over 45 evaluation reports on tens of thousands of data points, and improved average accuracy by ~22% after optimization
 - Constructed JSON templates for Q&A tasks including text summarization, sentiment analysis, information extraction, and created 2,000+ manually crafted prompts and answers for supervised fine-tuning
-

RESEARCH & PROJECTS

Fine-Grained Text Sentiment Analysis and Transfer | [report](#)

Sep. 2024 – Dec. 2024

- Generated a 3,000+ sentence emotion dataset across 6 emotions and 10 intensity levels using GPT and manual refinement
- Established a BERT-based emotion classifier with 98% validation accuracy by optimizing direct sentence processing
- Built a BERT regression model for emotion intensity estimation, achieving a 0.1 mean error rate with masking and scaling
- Engineered emotion transfer methods using Representation Engineering and T5 Transformer, validated via human evaluation

Causal Discovery Analysis of Remote Work and Mental Health Data | [report](#)

Sep. 2024 – Dec. 2024

- Led performance analysis of causal discovery algorithms (PC, FCI, GES) studying remote work's impact on mental health
- Implemented data simulations with linear and non-linear relationships between isolation, work hours, and stress variables
- Achieved 87.5% accuracy using Fisher Z, KCI, and ANM, validating a framework for analyzing remote work mental health

Bank Account Fraud Detection | [report](#)

Jan. 2024 – Mar. 2024

- Investigated Bank Account Fraud dataset, conducted EDA, statistical testing, feature selection, and Matplotlib visualizations
- Employed and optimized predictive models using Decision Tree, XGBoost, and CatBoost algorithms to detect fraudulent bank account applications, achieving 80% accuracy
- Conducted in-depth feature engineering and hyperparameter tuning using GridSearch with 5-fold cross-validation, leading to the identification of crucial features that contributed to enhanced model performance

Decades of Global Internet Usage: A Visual Overview

Jan. 2024 – Mar. 2024

- Developed the structure, style, and design of the [website](#) using Svelte, HTML, and JavaScript
- Crafted an interactive map visualization of global internet usage trends, integrating features like sliders, search functionality, ranking lists, and dynamic line plots for enhanced user interaction and detailed analysis