YUCHEN (ALICE) ZHANG

New York, NY | alicezhang727@gmail.com | (609) 751-8039

https://www.linkedin.com/in/yuchenalicezhang | https://github.com/zycalice

EDUCATION

University of Pennsylvania, GPA 4.0   **Philadelphia, PA/Remote |** September 2020 – May 2022 (Expected)

Master of Science in Engineering, Data Science

- Relevant Courses: Machine Learning, Stat for Data Science, Intro to Software Development, Computational Linguistics (NLP), Advanced ML, Big Data Analytics, Principles of Deep Learning (Intended)

Emory University, GPA 3.6   **Atlanta, GA |** August 2013 – May 2017

B.S. Quantitative Sciences, Political Science; B.B.A. Information Systems & Operations Management

- University of Edinburgh (Edinburgh, UK): Fall 2015 (Risk Management, Quality Management, Management of Technology)
- Relevant Courses: Statistical Learning, Time Series Analysis, Forecast & Predictive Analytics, Probability & Statistics, Advanced Statistics, Data Analytics & Visualization (+ TA), Game Theory

PROFESSIONAL & INDUSTRY EXPERIENCE

Claudius Legal Intelligence, Inc. (https://www.claudius.ai/)   **Remote |** August 2021 – Present

NLP Intern (Part time)

- Build Natural Language Processing models for legal applications

Momentive.ai (formerly SurveyMonkey)   **San Mateo/Remote |** June 2021 – August 2021

Data Science Intern, Machine Learning – Product Data Science Team

- Performed research and detected gibberish/irrelevant responses to open-ended questions using Sentence BERT, linguistics rules, and outlier detection methods; Improved f-score by 0.60; Conducted error analysis; Built demo app through Streamlit
- Hackathons: 1) Multilingual hate speech detection and 2) Bias in ML through adversarial model and sentence debiasing

Advanced Continuing Education Association   **Boston, MA/Remote |** June 2020 – August 2020

Data Science Intern

- Explored and improved course recommendation engines through research and **NLP** techniques (word2Vec, bio2Vec, TF-IDF)
- Built scrapping **pipelines** (Scrapy, Scprayd, Crontab, Git) for 3 spiders and scraped 1,000+ physician continuing educations courses (Selenium, BeautifulSoup) into SQL database
- Researched requirements and detailed regulations on continuing education to help **design data architecture**

NERA Economic Consulting   **New York City, NY |** July 2017 – June 2020

Analyst, Associate Analyst, Research Associate – Securities & Finance Practice

- Conducted quantitative and qualitative **research** to help produce defendable expert opinions in disputes and litigations involving insider trading, naked short selling, bond flipping, options trading, lottery fraud, student loans, and securities class actions
- Examined market and firm behaviors through **regressions and statistical tests** (parametric and non-parametric) (R, Python)
- Created **text analysis** (market sentiment/content analysis) on news and analyst reports for 5+ cases (R, Python, MS Excel)
- Analyzed, examined, wrangled, and manipulated **large and complex datasets** up to 10TB through statistics and visualizations to **find insights and detect trends** (Python, R, Stata, SAS, MS Excel)
- Estimated 10+ securities class action (section 10b-5, 11) damages up to billions of dollars using complex economic models
- Drafted expert reports and rebutted logical fallacies in opponent expert reports; communicated technical results to stakeholders

RESEARCH & PROJECTS

University of Pennsylvania, World Well-Being Project (WWBP)   **Remote |** January 2021 – Present

Research Assistant

- LifeSense: collaborate with psychologists (Northwestern) to understand text data with depression and stress
- Bias: examine and tackle classifier gender bias using n-gram features, LIWC, lexicon, and lexical weights

Ongoing Personal Project – Class Action Analysis Using Class Action Database (GitHub)

- Scrapped summary case table, case information, and downloaded documents under First Identified Complaint (FIC)
- Analyze the summary case table for recent trends and insights; produce ranking tables and graphs
- Predict if a case will be dismissed or settled with and without the FIC as features (Ongoing, best f-score without the FIC: 66.3%)

Ongoing Research Project – Gradual Domain Adaptation in NLP (Private Repository)

- Apply Gradual Domain Adaptation from CV to NLP using pseudo labels for text classification tasks; Used Amazon product reviews for different categories as dataset for sentiment analysis task

SKILLS

Python (scikit-learn, PyTorch, NLTK Tokenizer, Matplotlib), R (3-year work: tidyverse, dplyr, ggplot2), Scrapy, Scrapyd, Git, Linux/ Command Line, Jupyter, Colab, SQL, Stata, SAS (Basic), MS Office, FactSet, Bloomberg, Mandarin Chinese