Chelsea (Aiqi) Cui

(814) 882-9127 | ac4788@columbia.edu | acui34.github.io

EDUCATION

Columbia University New York, NY
  M.S. in Data Science Expected Dec 2021
Coursework: Applied Deep Learning, Cloud Computing & Big Data, Personalization

Carnegie Mellon University (with University Honor, GPA: 3.8/4.0) Pittsburgh, PA
  B.S. in Information Systems, Additional Major in Statistics & Machine Learning May 2020
Coursework: Machine Learning, NLP, Data Mining, Statistical Inference, Database, Data Visualization

SKILLS

- **Data Science Tools**: SQL (SQLite, PostgreSQL), D3, Spark, Hadoop, MapReduce, Tableau, Excel Pivot Tables
- **Programming Tools**: Python (NumPy, Scikit-Learn, Pandas, Keras, TensorFlow, NLTK, Matplotlib, ggplot), R, C
- **Software Development**: React, Ruby on Rails, Swift, HTML, CSS, TypeScript, Git, A/B Testing, AWS

WORK EXPERIENCE

Carnegie Mellon University, School of Computer Science, HCII Pittsburgh, PA
Research Assistant/Undergraduate Researcher Nov 2019 - Aug 2020
- Built a pipeline with Linux commands and Python to extract, clean, and reformat raw website log data
- Analyzed 10-day user interaction data through pattern recognition and clustering to model learning pathways
- Created an interactive system with D3 to visualize learning pattern differences between novice and professional users
- Collaborated with designers and converted data insights into feasible new features of an educational iOS app
- Submitted a paper to AERA 2021 for initial review as the first author

Handshake San Francisco, CA
Software Engineering Intern May 2019 - Aug 2019
- Performed A/B tests on search relevance and examined results with GCP and BigQuery to improve application rate
- Initiated company-wide web accessibility discussion on platform user interface leading to 10+ major changes
- Enhanced job role search performance and increased conversion rate by 262% in 7 days with Elasticsearch

Microsoft Guangzhou, CN
Software Engineering Intern May 2018 - Aug 2018
- Built an ML based classification filter with Python to raise chatbot's chitchat response accuracy from 70% to 96%
- Created and maintained a bilingual knowledge base with 3,000+ utterance data to support 95% business FAQs
- Analyzed weekly user responses via Excel Pivot Tables and Tableau to assess major feature releases

RELEVANT PROJECTS

NLP Auto Question Generation & Answering System, *CMU.* May 2020
- Designed and built a system to generate questions and answers automatically from Wiki articles as a team leader
- Applied techniques such as lemmatization, POS tagging, and NER to boost answering accuracy from 50% to 90%

House Prices Modeling and Prediction, *Kaggle Competition.* May 2018
- Conducted statistical EDA, imputed missing values, and explored multicollinearity to select key features
- Tested and evaluated multiple supervised models such as Random Forest, Lasso Regression and Gradient Boosting Machine and performed hyper-parameter tuning to predict house prices; Ranked top 15% worldwide

Cultural Difference Analysis on YouTube Trending Videos, *CMU.* April 2018
- Analyzed trending video data in three countries using Python Pandas library to derive area-specific business development insights for youtubers
- Presented analysis results and recommendations with innovative graphics (e.g. cloud map, heat map) via Tableau

Grubhub Database Reengineering, *CMU.* April 2017
- Redesigned conceptual and relational model of the Grubhub database to mock existing functions
- Developed the database with PostgreSQL with 10 complex queries to realize innovative use cases