

Yuchen Li

cornerliyc@gmail.com | (917) 756-1780 | www.linkedin.com/in/yuchen-li-profile

EDUCATION

Pennsylvania State University, College of Information Sciences and Technology

University Park, PA

B.S. Applied Data Science, Minor in Mathematics

December 2022

- Cumulative GPA: 3.6/4.0, Major GPA: 3.8/4.0
- Dean's List for 5 semesters
- Relevant Coursework: Machine Learning, Data Science Statistics, Big Data Analytics, Probability & Discrete Mathematics, etc.

PROFESSIONAL EXPERIENCES

Data Analyst, Holistic Solutions, LLC, Chicago, IL

January 2023 – September 2023

- Established and managed data pipelines, implementing tailored ETL solutions across diverse data sources (HRIS/ATS) with 5M+ observations by leveraging Google Cloud Platform (GCP) to generate automated reports and build dashboards.
- Created interactive dashboards via Looker Studio for 50+ clients to display multi-dimensional metrics (e.g. Moving Average), identifying opportunities for improvement in achieving DEI goals and crafting data-driven client narratives.
- Developed regression model on clients' Employment Engagement data to pinpoint correlations between 20+ factors such as perception of brands (e.g. Net Promoter Score) and benefits, resulting in 50% reduction in total analysis time.
- Partnered with client managers to support data requirements and managed multiple projects roadmap design, leading to the development of new measurement metrics and attracting resubscriptions from 10+ clients.

Data Science Research Assistant, College of IST, University Park, PA

May 2022 – December 2022

- Collaborated with 7 members to develop *Medical Q&A Chatbot Projects*, aimed to develop a medical dialog system using unsupervised learning; Leveraged knowledge graph for efficient patient FAQ responses.
- Employed web crawlers for data collection and established knowledge units by Named Entity Recognition (NER).
- Applied K-Means clustering to address entity polysemy and established entity-object connections based on Jaccard similarity.
- Compiled the executive summary outlining the strategy used and future roadmap, and presented results to the project's 4 executive members as the team leader.

Data Analyst Intern, CGL Consulting Co., Ltd, Shanghai, China

June 2021 – August 2021

- Completed the LinkedIn resume crawling project independently using Python incorporating various regular expression techniques and ultimately archived qualified resumes.
- Collaborated with the department manager to build a candidate transaction number model to forecast future number of candidates and assist candidate recruiting, reaching high accuracy.
- Performed daily data analysis of department KPI through data visualization and data wrangling using R for more than 10 branches regions.

SKILLS

- **Technical Skills:** Python, SQL, R, Excel, Google Cloud Platform (GCP), Business Intelligence Tools.
- **Big Data & Machine Learning:** Spark, Scala, Hadoop, MongoDB, Python (eg. scikit-learn, numpy, pandas, matplotlib)
- **Data Science & Miscellaneous Technologies:** Machine Learning, ETL, Data science pipeline (cleansing, wrangling, visualization, modeling, interpretation), Jupyter, Git, Mathematica, LaTeX
- **Certifications & Training:** Looker Studio Certification - *Google Analytics Academy*, January 2023

PROJECTS

Speech-to-Speech Translation, Project Team Member, University Park, PA

January 2022 – May 2022

- Collaborated on a multifaceted project with 2 members, covering Speech-to-Text, Text-to-Text, and Text-Speech components.
- Specialized in Text-to-Speech by retraining "YourTTS" multilingual zero-shot model with 3TB of audio clip data for Japanese and German, and fine-tuning on Embeddings, resulting in achieving a natural human-like timbre in the voice output.
- Authored the research paper and delivered a comprehensive presentation of projects.

The American Statistical Association (ASA) DataFest™, Project Team Member, University Park, PA

March 2022 – April 2022

- Utilized R to analyze in-game player data to identify optimization areas, including DEI and scene dwell time.
- Developed initiatives through visualizations and statistical analysis; Created data-driven storytelling to effectively communicate findings and offer actionable insights to optimize game settings and inform future development decisions.

Kaggle Competition – Stock Data Analysis and Prediction, Individual Project, University Park, PA *December 2020 – January 2021*

- Applied undersampling, preprocessing, and train/test split to process high-dimensional stocks datasets to predict stocks change and built binary classification model through KNN, GaussianNB, DecisionTree, RandomForest, and Ensemble methods.
- Achieved high F1 score of 0.46, ranked top 2 in the competition.