

Yiran Li

(919)-260-6753 | lyiran66@gmail.com | [linkedin.com/in/yiran-li626](https://www.linkedin.com/in/yiran-li626) | Emeryville, CA

EDUCATION

University of California at Berkeley
M.A. in Applied Statistics, GPA 3.57/4.0

Aug 2022 – Dec 2023
Berkeley, CA

University of North Carolina at Chapel Hill
B.S. in Statistics and Analytics, GPA 3.90/4.0, Summa Cum Laude

Aug 2019 – May 2022
Chapel Hill, NC

SKILLS

- **Programming:** Python (numpy, pandas, seaborn, scikit-learn, spark, tensorflow), SQL, R
- **Software:** AWS (S3, Lambda, Sagemaker), Databricks, Git, Power BI, Tableau, Microsoft (Excel, PowerPoint), JavaScript
- **Courses:** Experimental Design, Statistical Programming, Database Application, Machine Learning, Causal Inference, NLP
- **Skills:** Effective Communication, Analytical Expertise, Problem-Solving, Customer Engagement, Growth Mindset

PROFESSIONAL EXPERIENCE

FlexGen Power Systems, Inc.
Data Analyst Intern

Durham, NC
May 2023 – Aug 2023

- **Metric Design:** Developed Python scripts for RTE (Round-Trip Efficiency) calculations across 25 battery sites at diverse time intervals (annual/monthly/weekly), resulting in critical insights for site performance and optimal stability periods
- **Data Pipeline Automation:** Automated real-time data collection and analysis process by implementing RTE calculation logic via Databricks, achieving real-time insights every 5 minutes and a 2-hour reduction in runtime
- **Supervised Learning:** Trained random forest, XGBoost model of 50+ features to predict the battery SoC (State of Charge), resulting in essential insights for optimal BESS (Battery Energy Storage Systems) management
- **Data Visualization:** Created informative dashboards using Power BI on time-series data, presenting complex data insights in a user-friendly format, resulting in improved data communication and informed decision-making by stakeholders
- **Documentation:** Enriched the company's Confluence page by publishing a 16-page documentation on RTE methodology

University of North Carolina at Chapel Hill
Undergraduate Research Assistant

Chapel Hill, NC
Jan 2021 – May 2021

- **Data Preprocessing:** Transformed raw UK Biobank data including patients' age, gender, clinic records, lifestyle into an optimized model-ready format, slashing features from 664 to 226 and reducing model run time by 1 hour
- **Feature Engineering:** Enhanced model interpretability and performance by developing a composite "Lifestyle Score" by aggregating health factors (alcohol, smoking, diet, sleep, exercise) with weighted values assigned
- **Exploratory Data Analysis:** Initiated a thorough EDA process, leading a group of 3 to uncover significant patterns linking patients' diabetes status with their cholesterol, lipids, and blood pressure levels
- **Supervised Learning:** Employed rigorous model training and selection including LASSO, Random Forest and XGBoost, increasing the accuracy for diabetes detection by 20%

Fresenius Medical Care Investment (China) Co., Ltd.
Marketing Intern

Shanghai, China
Jan 2019 – April 2019

- **Data Visualization:** Boosted the average event ROI by 11.3% through quantitative analysis on events KPIs and visualizing key trends (locational, seasonal) in Tableau
- **Team Collaboration:** Led a collaborative effort between 2 stakeholders (third party vendors, internal functions) in event planning, resulting in 8 successful outcomes

HIGHLIGHTED PROJECTS

Predicting 2021 NFL Game Spread, Total and Result

- Aggregated, split, and rejoined different sources of game data (play-by-play, offensive/defensive stats, web scraped data) via SQL, building a structured dataset to be used for statistical modeling
- Engineered 1 metric to evaluate the outcomes for 2982 games, formulating sports betting strategy with 7% higher return
- Led and coordinated a team of 5 individuals, fostering dynamic brainstorming sessions and collaborative problem-solving, resulting in the team achieving the highest score within the class cohort

Detect the Negative Reviews from Customers

- Conducted text data pre-processing using techniques including stemming and lemmatization for it to be fed into the NLP model for further learning and sentiment analysis
- Improved user retention rate by 8% through pinpointing user pain points mentioned in reviews with topic modeling