CHLOE FENG https://www.linkedin.com/in/bixiao-feng-046b15304/
Email：fengbixiao857@gmail.com (619) 613-1963 https://github.com/rsm-b3feng

EDUCATION

Master of Science in Business Analytics 12/2025
University of California, San Diego
- GPA: 3.84/4.00 | Courses: ETL & SQL, Web Mining & Recommendation System, Scalable Analytics (Spark & AWS), Unstructured Data (NLP), Fraud Analysis

Bachelor of Science in Statistics 06/2024
The Chinese University of Hong Kong, School of Data Science, Shen Zhen, China
- GPA: 3.54/4.00 | Courses: Machine Learning, Statistical Inference, Stochastic Simulation, Data Structure, GLM
- Honors: 2021-2022 & 2022-2023 & 2023-2024 Dean's List, Outstanding Student Assistant

SPECIALIZED SKILLS

- **Machine Learning:** Regression, SVM, K-Means Cluster, Neural Networks, XGBoost, NLP (Word Embedding, IR)
- **Statistics inference:** Stochastic Simulation, Causal Inference, GLM, Bayesian Models, Causal Inference
- **Programming & Software:** SQL, Python (Pandas, Scikit-Learn, PyTorch, PySpark), R, PowerBI, Jupyter, AWS

EXPERIENCE

Data Scientist Intern, Dreamline AI, Orlando, FL 09/2024 - Present
- Utilized Hive SQL, PySpark, and API to extract and clean customer address data from 10G+ contract records, applying regular expressions for efficient data validation and ensuring data pipeline integrity.
- Analyzed California supplier data using Python and R, integrating regional user profiles to optimize expansion strategy, contributing to a $20,000+ cost saving.

Data Scientist Intern, Kuaishou Technology, Bei Jing, China 05/2023 - 09/2023
- Reduced effective cost per mille (eCPM) by optimizing creator payment policies using Hive SQL and Python, saving $200K and increasing video views by 10%. Evaluated incentive sensitivity to identify top revenue creators for operation teams.
- Detected 3K+ fraudulent videos using regression analysis on 300K+ records, leading to removal of 41 fraudulent creators and $10K cost savings.
- Designed a fraud-risk index and implemented an alert mechanism, automating daily identification of high-risk creators and preventing significant potential losses using internal BI system.
- Developed 4 auto-updated interactive dashboards using internal BI platform to analyze consumption patterns for 3 teams, enabling cross-departmental decision-making and saving 2 hours manual effort per day.

Data Scientist Intern, Ebao Technology Insurance (Shanghai) Co., Ltd, Shen Zhen, China 06/2022 - 09/2022
- Automated data pipeline (ELT) in Python by structuring 6 modular classes, processing 30GB+ of healthcare data cross multiple tables and reducing manual effort by 3 days per run.
- Conducted A/B testing for SMS marketing campaigns, optimizing recall times and SMS copy, leading to 12% promotion in purchase rates through SMS channel.
- Aggregated and cleaned 10M+ user records using SQL and Python, producing 6 region-specific reports with Tableau and Power BI to address unique insurance market demands and guide strategic planning.

PROJECTS

Leader, Fake Job Detection (NLP), San Diego 09/2024
- Cleaned, transformed, and explored 20K job postings using NLTK in Python, identifying key text features after removing stop words, tokenizing, and lemmatizing.
- Built frequency-based vector representations (BoW and TF-IDF) to convert processed text and developed classifiers (Logistic Regression, SGD, Random Forest, XGBoost), achieving F1-score of 0.82 and recall rate of 0.73.

Leader, Public Health-Covid-19 Analysis, Kaggle Competition 04/2023
- Led a team to study 3K+ training data to predict death of Covid-19 in Ohio. Visualized data by constructing histogram, conducted data cleaning through normalization, and utilized PCA for dimension reduction.
- Implemented neural network, random forest with cross-validation as predictive model, achieving R-squared of 0.89. Balanced optimal accuracy and avoided overfitting, ranking as top 1 among undergraduate students.
- Project URL: https://www.kaggle.com/competitions/covid19casesinohio-spring2023