Qifei Cui

55 McDonald Pl, Goleta, CA 93117

📞 805-971-1388 ✉ cqfsq@outlook.com 🔗 linkedin.com/in/qifei-c-339a76247 �नhttps://github.com/Qifei-C

Education

University of California - Santa Barbara   **Sep. 2019 – Dec 2023**
B.S. in Applied Mathematics   *Goleta, CA*
Dual - B.S. in Applied Statistics and Data Science

GPA: 3.74/4.00 *(Upper Division Courses in Data Science 3.95/4.00)*

Relevant Coursework

- Advance Linear Algebra (A)
- Mathematical Statistic (A)
- Ordinary Differential Equation (A)
- Partial Differential Equation (A)
- Numerical Analysis (A)
- Bayes Analysis (A)
- Time Series (A)
- Stochastic Process (A)

Experience

Research in Sequential Recommendation System   **Aug 2023 – Present**
Supervised by Professor Haowen   *China*

- Pioneering a solution to the cold start problem in sequential recommendation systems by incorporating 20-dimensional embeddings for both users and items, derived from statistical properties.
- Enhancing the base BERT4Rec model with dynamic graph layers, inspired by Graph Convolutional Networks, to improve recommendation accuracy and adaptability in user-item interactions.
- Conducted experiments on the IMDb1M dataset, diverging from conventional recommendation system approaches. Testing sets are strategically selected, focusing on the 30% of users with a watching history of less than 50 titles, to address diverse user behaviors.

Capstone Project   **Sep 2023 – Jan 2024**
Supervised by Professor Laura   *Santa Barbara, CA*

- Collaborative Project: Collaborated on a project exploring data science applications, using GitHub for version control and team coordination.
- Applied in Real-World Circumstances: Utilized tidyverse for in-depth analysis of survey data, encompassing hypothesis formulation and classification model development. Implemented keras to enhance web fraud detection accuracy, focusing on text processing. Directed soil temperature data analysis with time series and spatial techniques.
- Teaching and Vignette Development: Participated in a teaching exchange, developing and workshopping educational vignettes.

Alphabet, Inc   **Mar 2023 – June 2023**
Intern Data Analyst   *CA*

- **Recommendation System:** Implemented efficient pipelines for preprocessing large datasets like Anime, Amazon, and IMDb, ensuring data integrity. Developed a sophisticated double-tower recommendation model adapted from recent Recommendation System research papers. Utilized user ID input to rapidly generate tailored movie suggestions with detailed information.
- **Embedding Construction:** Engineered new embeddings by analyzing user-item interaction data based on statistical distributions. Applied different levels to varying lengths of watching history, resulting in a 2.3% increase in final recommendation accuracy.
- **Model Evaluation:** Refined the model using k-fold cross-validation on a dataset of 100k users and 2 million movies, achieving a accuracy of 0.793 on Recall10, with validation loss of 1.3150 and a training loss of 1.2490, demonstrating effective learning and robust generalization.

United World College - Changshu China   **Sep 2021 – June 2022**
Intern Teacher in Math/Science Dept   *Changshu, Suzhou*

- **Teaching Outcomes:** Tailored instruction to address students' neglected, weaker knowledge areas, guided by their academic history and performance. Sstudents achieved an average 20% improvement in simulated IB math exam scores, with the highest increase being from 1 to 6 points (on the IB 7 grading scale).
- **Internal Assessment Advisor**: Supervised two student projects in Mathematics.
- **Student Activity Supervisor:** UWC student project "Zhixing - Listen to LGBTQ+" supervisor, supervised and advised on project progress.

Projects

Monthly Electric Power Generation in the U.S. | *R, Time Series, SARIMA* **Dec 2023**
- **Data Analysis:** Analyzed electricity production data from January 1985 to January 2018. Performed data preprocessing including dealing with non-stationarity and heteroscedasticity, and implemented a training and testing split for model validation.
- **Model Selection and Refinement:** Employed SARIMA model for time series forecasting, Utilized techniques like differencing and logarithmic transformation for variance stabilization and trend neutralization. Selected candidate model through conducting a thorough evaluation using ACF, PACF, and Periodogram.
- **Model Evaluation:** Conducted a thorough evaluation of residual of the model with statistical evidence that the residual follows gaussian white noise.

Risk User Identification | *Python, R, Deep Learning, Random Forest, PCA, SMOTE, Stratified Sampling* **July 2023**
- **Data Processing:** Cleaned and preprocessed a dataset comprising 40k user records include replaced missing values and outliers, encoded 43 distinct features, and standardized data. Utilized PCA to reduce dimensionality to 12, while keeping 95% of variance. Addressed SMOTE to generate synthetic samples and achieve a balanced class distribution.
- **Model Development:** Built a multi-layer deep learning model using Keras, achieving an initial accuracy of 84.7%. Conducted 10-fold cross-validation with stratified sampling, achieving a mean accuracy of 83.2%. Tested the integrated model on a separate holdout set, confirming a final accuracy of 74.3%.
- **Model Improvement:** Integrated a Random Forest classifier to further filter potential false positives and false negatives. Same cross-validation applied and the RF model achieved a mean accuracy of 78%. Overfitting in the deep learning model was addressed using dropout layers and early stopping. The final model has a mean accuracy of 80.15% and an accuracy of 79.8% on the separate set.

Fraud Transaction Detect | *Python, Deep Learning, SMOTE* **May 2023**
- Developed a deep learning model to detect fraudulent transactions using Python and popular deep learning libraries.
- **Data Processing:** Utilized a dataset of over 100,000 transactions, where 0.2% represented fraudulent activities. Employed the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples, ensuring a balanced training set.
- **Model Development:** Incorporated various features, including transaction amount, location-based data, transaction time , and 29 anonymous features.
- **Model Outcome:** Achieved a precision of 89.4% and a recall of 88.2%, significantly improving the detection rate of fraudulent transactions compared to classic logistic regression methods.

Numerical Analysis Tool Package | *Python, Matrix Analysis, ODE, PDE* **Jan 2023**
- Incorporated matrix analysis tools, enabling efficient computation of eigenvalues, eigenvectors, matrix inversion, decomposition and factorization. Developed solvers for ODE using Runge-Kutta and Euler's method. Integrated algorithms for Partial Differential Equations (PDE), suitable for numerical analysis on 1D and 2D heat equation.

Technical Skills

Languages: Python, R, SQL, SAS
Developer Tools: VS Code, Jupyter Notebook, Rstudio, Google Cloud Platform
Technologies/Frameworks: GitHub, Anaconda, Node.js, Pytorch, Tensorflow, Tidyverse